Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
WALTER S. MONROE

VOL. IV JUNE, 1921 No. 1

CONTENTS

DIAGNOSTIC AND REMEDIAL STEPS IN READING..... W. S. Gray 1
THE NATIONAL INTELLIGENCE TESTS..... G. M. Whipple 16
THE MEASUREMENT OF LANGUAGE: WHAT IS MEASURED AND ITS SIGNIFICANCE..... E. J. Ashbaugh 32
INTELLIGENCE TESTS IN CLASSIFYING CHILDREN IN THE ELEMENTARY SCHOOL..... Charles Fordyce 40
THE INTELLIGENCE EXAMINATION FOR HIGH-SCHOOL FRESHMEN..... Ira J. Bright 44
EDITORIALS 56
REVIEWS AND ABSTRACTS..... 61
NEWS ITEMS AND COMMUNICATIONS..... 68
NATIONAL DIRECTORS OF EDUCATIONAL RESEARCH... 75

PUBLISHED FOR THE BUREAU OF EDUCATIONAL RESEARCH • COLLEGE OF EDUCATION • UNIVERSITY OF ILLINOIS •

Journal of Educational Researc

Copyright 1921, by
Public School Publishing Co.

Published for the Bureau of Educational Research, University of Illinois

Editor-in-Chief: **B. R. Buckingham**
Director of Bureau of Educational Research, University of Illinois

Associate Editors

E. J. ASHBAUGH, *Ex-Officio, Secretary of Research Association, University of Iowa, Iowa City.*
W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*
S. A. COURTIS, *City Schools, Detroit, Mich.*
WALTER S. MONROE, *University of Illinois, Urbana, Ill.*
GEORGE D. STRAYER, *Columbia University, New York City.*
LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, Calif.*
GUY M. WHIPPLE, *Director, Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*
CHARLES FORDYCE, *Dean Teachers' College, and Director of Bureau of Educational Measurements, University of Nebraska.*
WILLIAM S. GRAY, *Dean College of Education, University of Chicago.*
M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota.*
V. A. C. HENMON, *Director School of Education, University of Wisconsin.*
ARTHUR W. KALLOM, *Assistant Director Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*
F. J. KELLY, *Dean, School of Education and Director Bureau of School Service, University of Kansas.*
ROBT. H. LANE, *Assistant Superintendent of Schools, Los Angeles, Calif.*
WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*
GEORGE W. MELCHER, *Director Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*
EUGENE A. NIFENECKER, *Assistant Director Bureau of Reference, Research, and Statistics, Public Schools, New York City.*
JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director Efficiency Bureau, Public Schools, Rochester, New York.*
LEON O. SMITH, *Assistant Superintendent (in charge of Education Research) Public Schools, Omaha, Nebraska.*
W. W. THEISEN, *Director Division of Reference and Research, Public Schools, Cleveland, Ohio.*
CLIFFORD WOODY, *Professor of Education, University of Washington.*

Editor of "Reviews and Abstracts"

E. H. CAMERON, *University of Illinois, Urbana, Ill.*

THE JOURNAL OF EDUCATIONAL RESEARCH is published at 450 Ahnaip St., Menasha, Wis., monthly except July and August by Geo. Banta, Official Printer to the Public School Publishing Co. The subscription price is $4.00; $4.20 in Canada or Mexico; and $4.30 in other foreign countries.

The editorial office is at Urbana, Illinois. Address all communications on educational matters, articles for the magazine, etc. to B. R. Buckingham, University of Illinois, Urbana, Illinois.

The Public School Publishing Company conducts the subscription work. Send all orders to them at Bloomington, Illinois.

The JOURNAL OF EDUCATIONAL RESEARCH is both needed by and needs the subscriptions of many more teachers and school officers.

The title page and index for Volume III were furnished with May, 1921 number.

Articles to Appear Soon

The New Plan of Admitting Students at Columbia University by *Edward L. Thorndike*.
Analysis of Reading Ability by *S. A. Courtis*.
Measuring Progress by Means of Standardized Tests by *S. S. Brooks*.
Arithmetic Ability of Men in the Army and Children in the Public Schools by *Arthur Kolstad*.
Correlation: How to Work It on the Adding Machine by *Herbert A. Toops*.
Rate of Progress in Teacher Preparation by *W. Randolph Burgess*.
Method of Equalizing Teachers' Ratings by *Lee Byrne*.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV JUNE, 1921 NUMBER 1

DIAGNOSTIC AND REMEDIAL STEPS IN READING[1]

WILLIAM S. GRAY
The University of Chicago

Any report which is made at this time of diagnostic and remedial work in reading must be tentative and incomplete for two reasons. The technic of diagnosing individual and group needs has not been fully developed. Furthermore, appropriate remedial devices have been organized for only a limited number of disabilities. It is possible, on the other hand, to secure helpful suggestions from a study of specific methods now employed in various laboratories and school systems of the country.

Experience has proved that diagnostic and remedial work in reading should begin with a study of the accomplishments of all the pupils of a school. This step enables teachers to determine the major problems of reading instruction in each grade for the year. It also secures information concerning the needs of every pupil. Unless a systematic program of testing is followed there is danger that many significant individual difficulties will not be noted.

In the Elementary School of the University of Chicago at least four records of accomplishment in reading are secured at the beginning of each school year: (a) mastery of the rudiments, or mechanics of oral reading, including both rate and accuracy, as measured by the Standardized Oral Reading Paragraphs; (b) rate of reading simple material silently as measured by the Courtis' Silent Reading Test, No. 2; (c) ability to understand simple passages as measured by Courtis' Silent Reading Test, or Burgess' Scale for Measuring Ability in Silent Reading (P. S.-1); (d) ability to understand increasingly difficult passages, as measured by Monroe's Silent Reading Test, or Thorndike's Scale Alpha 2. Although these tests do not cover all phases of reading, they secure a

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

sufficient amount of information for an exceedingly helpful preliminary diagnosis. In the report which follows the discussions of diagnostic and remedial steps are limited on account of time to one phase of reading.

As soon as the tests have been given and the scores calculated, a careful study is made of the average scores of each grade to determine which phases of reading should be emphasized most during the year. For illustration, Figure 1 represents the average scores in oral reading in October, 1919. The vertical lines are the lines on which the scores for the grades are indicated. The solid oblique line represents the standard scores. The dotted oblique line represents the scores for the University Elementary School in October, 1919.

An analysis of the records revealed two significant facts: (a) oral reading needed little emphasis in each grade as a whole above the III-B, because the average class scores at the beginning of the year were higher than the standard scores for the end of the year; and (b) more or less class instruction in oral reading was necessary in the II-B, II-A and III-B grades because the average class scores were lower than the standard scores.

The second step in the diagnosis included an analysis of the individual scores in each grade. Table I shows the distribution of the pupils of each grade in oral reading. The numbers above the vertical columns indicate the oral reading scores. The grades are indicated in the left-hand column and the total number of pupils in each grade is indicated in the right-hand column. The entries in the table show the number of pupils receiving each score. Thus in the II-B grade, 1 pupil made a score of 5, 1 a score of 10, 3 a score of 20, etc.

The irregular vertical line in the table indicates the scores which pupils should make in oral reading to do most effective work in fluent intelligent reading. In the University Elementary School, these scores are 40 for the II-B grade, 45 for the II-A grade, 50 for the III-B grade, and 55 for all of the higher grades. The scores were determined three years ago after careful comparative studies had been made of the accomplishments of pupils in various phases of reading.

The pupils in the lower grades who ranked higher than these scores were promoted to a higher grade, if their work in other subjects justified promotion. If they were not promoted, they were

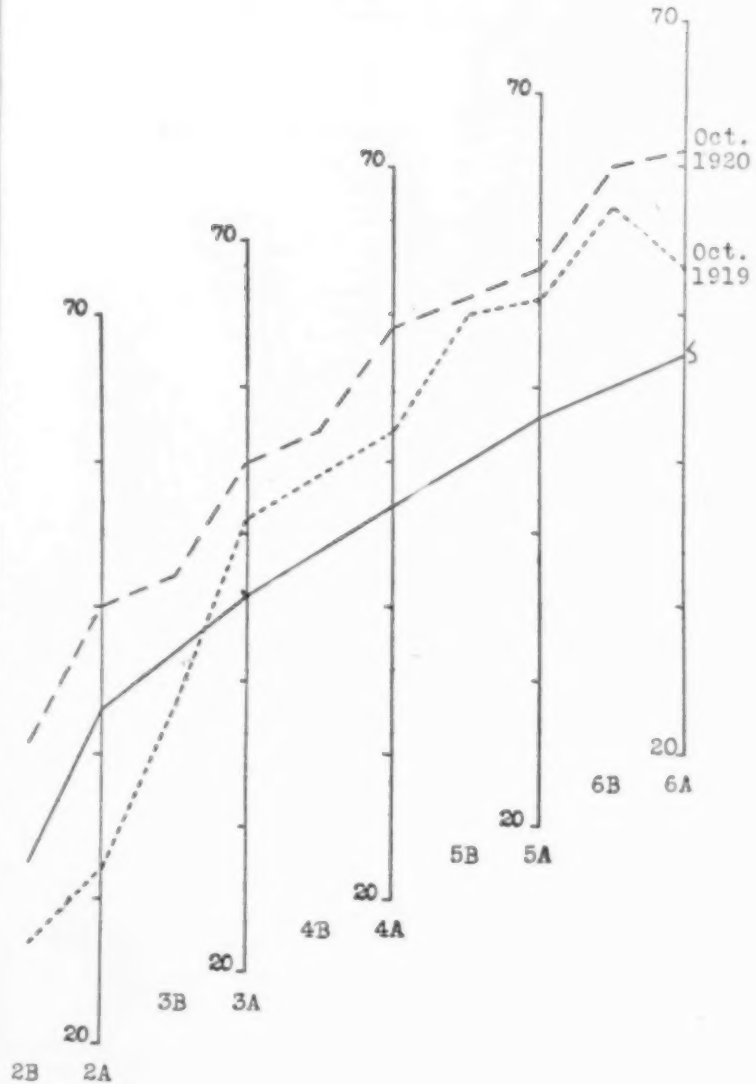

FIGURE 1. AVERAGE SCORES IN ORAL READING IN OCTOBER, 1919, AND OCTOBER, 1920

excused from oral-reading exercises, and either provided with material to read silently or given additional work in those subjects in which they needed help. The pupils who scored lower than the desirable standards were given help in oral reading in proportion to their needs. In attempting to organize appropriate remedial instruction for these pupils, additional diagnostic steps were adopted.

TABLE I. DISTRIBUTION OF PUPILS IN ORAL READING

Grade	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	Total
II-B	1	1		3	1	2	2	3	4	2	5	5		1		30
II-A				1		1		2	3	4	7	4	2	2		26
III-B							3	1	6	8	4	3				25
III-A									3	5	8	4	6		2	28
IV-B								3	8	3	8	3	4		3	32
IV-A										5	5	8	8	3	1	30
V-B								2	1	2	3	7	6	3		24
V-A									1	2	3	9	10	5	1	31
VI-B									1	2	4	9	6	6		28
VI-A										2	4	2	13	5		26

A study was first made of the kinds of errors which pupils make in reading. The most significant errors with words which were discovered are non-recognitions, total mispronunciations, and partial mispronunciations of (a) monosyllabic words and (b) polysyllabic words. The most important errors with groups of words are poor phrasing, omissions, insertions, substitutions, repetitions, and reversing the order of words and phrases.

A careful study was next made by the teachers of the methods of instruction which are appropriate in preventing or eliminating each type of error. Such matters were emphasized as the following: the importance of numerous simple interesting reading exercises

to promote the establishment of effective habits of reading; the value of flash-card exercises in the lower grades in developing a sight vocabulary, and in increasing the span of recognition; the importance of directing attention to the content of what is read in correcting mispronunciations; the value of phonetic analysis in eliminating partial mispronunciations of monosyllabic words; the value of a study of the rules of syllabication and accent in eliminating the mispronunciation of polysyllabic words; the importance of exercises in recognizing thought units in sentences in order to improve phrasing; and the value of careful reading to eliminate omissions, substitutions, insertions, and repetitions.

KINDS OF ERRORS IN ORAL READING

I. With Words
   1. Non-recognition
   2. Total mispronunciation
   3. Partial mispronunciations
      a. Monosyllabic words
         Beginnings and endings of words
         (Usually a consonant or phonogram)
         Middle parts of words
         (Usually a vowel or digraph)
         Enunciation
      b. Polysyllabic words
         Syllabication
         Accent
         Repetition of one or more parts
         Substitution
         Omission or insertion
         Inaccurate pronunciation of a part or a syllable

II. With Groups of Words
   1. Poor phrasing
   2. Omissions, insertions, and substitutions
      a. Which change meanings
      b. Which do not change meanings
   3. Repetitions
      a. To secure a better attack on a word or phrase
      b. To correct an error in pronunciation
      c. To verify meanings
   4. Reversing the order of words and phrases

The study of devices for overcoming difficulties was accompanied by a detailed analysis of the causes of difficulty in oral reading. Some of the more important causes which have been discovered

appear in the outline. Irregular attendance causes slow progress on the part of some pupils. School physicians and teachers frequently find that poor health, malnutrition, or nervous disorders interfere with rapid progress. Studies of the nationality of the child, his home environment and training, the economic conditions under which he lives and the peculiarities of the section of the country from which he comes have suggested remedial measures, such as a larger amount of training in English to overcome language handicaps, or a generous provision of interesting reading material which may be read at home. Furthermore, studies of the previous instruction of the child have revealed the fact that his difficulties are due to inappropriate or ineffective methods of instruction, to constant use of difficult selections, to an inadequate amount of reading material, or to uninteresting selections which were used.

The organic causes of difficulties in reading are far more difficult to detect and to remedy. Many children experience genuine difficulty because of visual defects which can be remedied through the use of appropriate lenses. Some children find it difficult to form the motor coordinations which are essential in reading. Carefully prepared exercises to strengthen the muscles of the eyes have proved helpful. A limited visual span prevents some readers from recognizing a group of words at a given fixation. This limitation can be detected through the use of a tachistoscope. Short exposure exercises have proved effective in increasing the amount which one can recognize at a single fixation of the eye. A low degree of visual acuity frequently leads to numerous errors in reading. It can be detected through a letter-marking test, and has been improved by exercises of the same type.

Defects in the brain tissues frequently cause difficulties in reading. Word blindness, for illustration, is defined as extreme difficulty in learning to recognize printed or written language by persons of normal mentality and vision. It has been overcome in a limited number of cases by methods of instruction which make a vigorous appeal to the child. Vocal defects, such as malformations of the vocal organs and enlarged tonsils, interfere seriously with effective reading. Breathing irregularities, which are caused by adenoids or by inadequate control of the diaphragm and chest muscles, may lead to errors. Auditory defects, such as word or sound deafness, frequently lead to errors in pronunciation.

A third group of causes are psychological in character. They include general mental incapacity, inadequate attention to meaning, failure to associate appropriate meanings with words, limited eye-voice span, limited span of recognition, inability to remember new words easily, capacity to learn words only very slowly forgetting them quickly and easily, and inability to analyze and pronounce words effectively.

With the various causes of difficulty in mind, the teachers made a careful study of the records of their pupils. Those who were below the required standard but who gave evidence of no unusual or peculiar difficulty were given oral-reading instruction appropriate for normal second- and third-grade classes. Special effort was made, however, to interest the pupils and to provide them with a large amount of simple interesting material for use in class and for supplementary outside reading.

The regular instruction which has just been described was supplemented by special exercises to meet the needs of small groups of pupils. For illustration, one group did considerable silent reading to develop habits of thoughtful reading. A second group gave attention to word analysis and phonics to gain independence in word recognition. A third group drilled on quick perception exercises to increase the span of recognition. A fourth group selected thought units in sentences to improve phrasing. Although much time was devoted to these exercises, the pupils were always given abundant opportunity to read in order that the help which was derived from the special exercise might be reflected as soon as possible in their reading habits.

The pupils who did not respond to this treatment were sent to the special "remedial" teacher for a thorough diagnosis and for appropriate remedial exercises. In order to describe concretely the diagnostic and remedial steps which were employed, the case of a fourth-grade boy will be presented in detail, who found it necessary in September, 1920, to give up some of his school work because of inability to read. This situation was so acute that it seemed for a time that he would have to discontinue school altogether. In the diagnosis of his case he was first given the Standardized Oral Reading Test in which he scored distinctly below the standard for the second grade. His record showed that he was unable to recognize many very simple words. He made

R.H.S

numerous substitutions, such as *says*, for *said*, *was* for *were*, *small* for *same*, and *he* for *the boy*. Furthermore, he recognized words individually rather than in groups, which resulted in an unusually slow rate of reading.

The most characteristic difficulty which was revealed was a peculiar form of confusion which led to frequent repetitions. In this connection he frequently pronounced words which appeared to the right before pronouncing those to the left. This error occurred so frequently that detailed consideration was given to it later in the diagnosis.

The Burgess' Scale for Measuring Ability in Silent Reading (P. S. -1) was given next to determine how effectively the subject read silently. In this test he ranked distinctly below the median score for the second grade.

Courtis' Silent Reading Test was given next to secure additional information concerning the subject's rate and comprehension in silent reading. In accuracy of interpretation he excelled the fourth-grade standard, but in rate of reading he ranked very low. The results of this test suggested quite definitely that his major difficulty related to the recognition of words, rather than to their interpretation.

Additional evidence supporting this conclusion was secured through a supplementary test based on passages of the Courtis rate test. After the subject had read three passages silently, he was asked to reproduce what he had read. He reproduced only 16 words, or 12.5 percent of the amount read. He then read three passages orally and reproduced 18 percent. Three passages were then read to him and he reproduced 70.1 percent of the content. The superior score which was made when the subject was relieved of the problem of recognition made it clear that his difficulty in reading related primarily to recognition rather than interpretation. A study was therefore made of the causes of difficulty in recognition. In this connection four steps were taken.

1. Jones' Vocabulary Test was first given to determine the ability of the subject to pronounce words accurately. The records showed that he failed to see many letters unless his attention was called to them, he confused the sounds of important letters of words, he was unable to analyze short words containing the simplest phonetic elements, and was unable to recognize at sight frequently recurring words, such as *what*, *that*, and *you*.

2. The difficulty encountered in the recognition of words suggested the possibility of marked weakness in visual memory. A visual memory exercise was therefore given to the subject and to four fourth-grade pupils who ranked distinctly above the average in general intelligence, school marks, and oral reading accomplishment. The special subject of this study made fewer errors than any other pupils. This indicated that very little if any explanation for slow progress in reading could be attributed to gross defects in visual memory.

3. A study was next made of the ability of the subject to recognize individual letters, parts of words, and groups of words at single fixations of the eye. The materials which were used included the following: the 26 letters of the alphabet, 18 two-letter words, and 10 each of two-letter nonsense syllables, four-letter words, four-letter nonsense syllables, two-word phrases, and three-word sentences. These were presented uniformly through the use of a drop tachistoscope. Each item was presented until it was accurately recognized. The four fourth-grade pupils who took the visual memory test took this series of tests also. Table II contains a summary of the results.

TABLE II. THE NUMBER OF ERRORS IN RECOGNITION BY FIVE SUBJECTS IN THE TACHISTOSCOPIC EXPERIMENT

Recognition Series	Subject A	Subject B	Subject C	Subject D	Subject E
Isolated letters	13	16	15	9	20
Two-letter words	5	6	10	1	2
Two-letter nonsense syllables	7	4	2	1	13
Three-letter words	10	4	4	0	3
Three-letter nonsense syllables	31	9	12	2	6
Four-letter words	19	1	1	2	4
Four-letter nonsense syllables	38	7	17	2	9
Two-word phrases	32	3	4	1	2
Three-word sentences	25	3	16	1	5
Total	180	53	81	19	64

Subject A made more than twice as many errors as any other subject. He recognized isolated letters, two-letter words, and two-letter nonsense syllables as accurately as the other subjects. His difficulties increased markedly with the three-letter units. The

numerous errors made in recognizing three- and four-letter nonsense syllables indicated marked inferiority in the accurate recognition of the details of a group of letters. Furthermore, the errors made in recognizing two- and three-word phrases and sentences indicated that his span of recognition was much narrower than that of the other subjects.

4. Before drawing final conclusions a photographic record was made of the eye-movements of the subject while reading a simple, unfamiliar passage silently. A summary of the difficulties which were discovered follows. (a) The subject recognized each word individually. This is indicated by the fact that there is at least one fixation, and in some cases, several fixations, per word. (b) These fixations were very irregular and unsystematic. This indicates that the subject did not have a definite method of extricating himself from a word difficulty. (c) The subject used an inaccurate, uneconomical return sweep from the end of one line to the beginning of the next. This was indicated by the location of the first fixation of each line. In the second line it was fourth from the left, in the third line it was eighth, in the fourth line it was twelfth, and in the fifth line it was seventh. (d) Instead of beginning at the left and going forward step by step, the eye skipped about, sometimes fixating on a point very much ahead of where it should be and at other times moving to the left over parts which had already been read. A partial explanation for this irregularity is found in a statement made by the subject. He stated that in reading a sentence he tried to find a sufficient number of words which he knew to enable him to guess at the meaning of the rest. This resulted in irregular wandering movements rather than definite progressive movements and made fluent reading more or less impossible.

The diagnosis which has been reported thus far revealed five outstanding defects in the reading habits of the subject: (a) inappropriate motor habits in making the return sweep; (b) irregular progression of attention from left to right; (c) failure or inability to scrutinize words in sufficient detail to recognize significant parts; (d) inability to analyze new words; and (e) inability to recognize words in groups or thought units.

The remedial program of instruction which was adopted included three distinct features. For thirty minutes each day the subject was under the immediate supervision of Miss Delia Kibbe, a special teacher of unusually bright and slow children in the Uni-

versity Elementary School. The problem assigned to Miss Kibbe was the development of ability to recognize words independently. During a second period of thirty minutes the subject was under the direction of the writer. An attempt was made (a) to correct the inappropriate motor habits which were employed in making the return sweep; (b) to promote regular progression of attention from left to right; (c) to increase the ability of the subject to note important details of words; and (d) to increase the span of recognition.

In order to develop effective eye-movements interesting selections were typewritten in three forms. In each type the lines were separated more widely than in ordinary print in order to promote a rapid and accurate return sweep. The subject was instructed to move his eyes quickly from the end of one line to the beginning of the next without stopping to look at any of the words of the second line. After ten five-minute exercises in which this problem was emphasized, he was able to make the return sweep effectively.

The words of the three types of exercises were typewritten in modified form in order to promote regular progression of attention from left to right. In the first type the words were written five-letter spaces apart. The subject was instructed to read the words in order without glancing to the right or to the left of a given word before it was recognized. The fact that the words were separated made it somewhat easier for him to concentrate attention effectively. After ten five-minute exercises of this type he showed marked improvement in regularity of eye-movements and in fluency in reading.

In the second type of exercise the words were grouped together in thought units as far as possible. Exercises of this type are calculated to promote a regular progression of eye-movements, and in addition, the recognition of more than one word at each fixation. If, while reading, the subject gave evidence of wandering movements, his attention was directed to the specific word or words which were causing difficulty. It should be added at this point that during a period of ten exercises the number of corrections which were necessary gradually decreased.

In the third type of exercise the words were written one-letter space apart. The subject read aloud five minutes each day while the experimenter noted for evidences of ineffective return sweeps and irregular progression of attention. If irregularities were

DRILL PASSAGES TO ESTABLISH APPROPRIATE EYE-MOVEMENTS

Type I

Once upon a time there lived in a cottage near a wood a poor widow. In the garden in front of her house there grew two rosebushes, one of which bore white roses and the other red.

Type II

One day when her spindle was so red with blood that the poor girl could not spin, she tried to wash it in the water of the spring; but the spindle fell out of her hand and sank to the bottom.

Type III

These two little girls were the best children in the world. Snow-White was quiet and gentle. She used to stay at home with her mother, help her about the house-work, and read to her after it was done; while Rose-Red liked to run about the fields and look for birds and flowers.

noted, suggestions, calculated to correct the difficulty, were offered. At the end of nine weeks marked improvement had been made as evidenced by greater fluency and fewer irregularities.

A second series of exercises was organized as a means of directing the attention of the subject to important details of words and of increasing the span of recognition. These exercises took the form of eleven drill books each of which contained ten words or phrases of approximately the same length for use in short exposure exercises. The first exercise of each book appears in Table III. The words of Book I were exposed in order. Less than a second was allowed for each exposure. A score of one was allowed for each word accurately recognized. The table shows that on the first day the words of the first three books were recognized with-

TABLE III. RECORD OF THE SHORT EXPOSURE EXERCISES DURING A PERIOD OF TWELVE DAYS

Book	First Exercise of Each Book	Day	Score Made at Each Exposure	Number of Exposures
I	on	1	10	1
II	has	1	10	1
III	bank	1	10	1
IV	of wind	1	6, 8	9
		2	8, 8, 6, 6, 10, 10	
		3	10	
V	in the garden	3	7, 7, 7, 7, 7, 9, 9, 10, 8, 10, 10	12
		4	10	
VI	the willow buds	4	7, 6, 6, 6, 6, 8, 7, 10, 10	17
		5	6, 9, 9, 9, 9, 10, 10	
		6	10	
VII	he said	6	9, 9, 9, 9, 10, 10	7
		7	10	
VIII	pretty soon	7	9, 9, 10, 10	5
		8	10	
IX	What is that?	8	3, 3, 5, 6, 6, 7, 8, 9, 9, 9, 10, 10	13
		9	10	
X	to her fairy story	9	2, 2, 4, 6, 7, 7, 8, 9, 8, 10, 10	19
		10	7, 9, 9, 10, 9, 10, 10	
		11	10	
XI	their little seed boxes	11	3, 3, 2, 5, 5, 6, 7, 10, 9, 10, 10	15
		12	8, 10, 10	
		13	10	

out errors. The phrases of Book IV were exposed twice on that day. During the first trial six of the ten phrases were accurately recognized. The subject was given his score but no statement was made concerning the character of his errors. During the second trial he made a score of 8. On the second day the exposures were continued until the subject had recognized all of the phrases without error twice in succession. On the third day the phrases were exposed again. Inasmuch as the subject made a perfect score on Book IV, Book V was begun.

The procedure which has just been described was continued for twelve days. At the end of that time the tachistoscopic exercises which had been given in the early examination of the subject were repeated. The table includes a summary of the errors made by the subject before and after training on the twelve daily exercises. It shows a decrease in the total number of errors from 180 to 122. There was a decrease in non-recognition from 39 to 32, and in wrong recognitions from 141 to 90. An examination of the records for each type of material used shows satisfactory improvement in the recognition of isolated letters, three-letter words, three-letter nonsense syllables, four-letter words, two-word

TABLE IV. THE NUMBER OF ERRORS IN RECOGNITION BY SUBJECT "A" IN THE TACHISTOSCOPIC EXPERIMENT BEFORE AND AFTER TWELVE SHORT EXPOSURE EXERCISES

Recognition Series	No.	Number of Errors Before Training			Number of Errors After Training		
		Non-recognition	Wrong Recognition	Total	Non-Recognition	Wrong Recognition	Total
Isolated letters	26	6	7	13	4	0	4
Two-letter words	18	0	5	5	3	2	5
Two-letter nonsense syllables	10	1	6	7	4	2	6
Three-letter words	10	0	10	10	2	2	4
Three-letter nonsense syllables	10	4	27	31	3	12	15
Four-letter words	10	4	15	19	1	1	2
Four-letter nonsense syllables	10	8	30	38	13	42	55
Two-word phrases	10	6	26	32	0	15	15
Three-word sentences	10	10	15	25	2	14	16
Total		39	141	180	32	90	122

phrases and three-word sentences. For some unexplained reason the subject did less well on four-letter nonsense syllables than in the original test. Inasmuch as the number of errors made in the recognition of several of the exercises was distinctly above the average made by effective readers in the first test, it was evident that short exposure exercises could be continued to advantage.

After two months of remedial work, the subject was thoroughly tested again to determine progress. In the oral reading test, he made a score of 37.75. This indicates that he made approximately a year's growth between October and December.

In the Burgess test, he made a score of 50 on the third-grade scale and a score of 34 on the fourth-grade scale. The satisfactory improvement may be attributed almost wholly to increased rate of reading.

In the Courtis' Silent Reading Test the subject answered twenty questions which is twice the number answered in the first test. He also made a comprehension score of 90 which is equal to the score in the first test.

The December tests showed clearly that the remedial work had resulted in greater fluency and effectiveness in reading. Although the subject was not fully up to standard by the middle of the current school year, he was permitted to resume regular work with the fourth-grade class, and he promises at the present time to make a very creditable record. If he had not been subjected to individual diagnosis, it is probable that the real nature of his difficulty would not have been discovered. Too many pupils fall into the retarded group, and finally drop out of school, because of difficulties which could be removed. In my judgment, each school system of considerable size should establish a center to which unusual cases could be taken for diagnosis and remedial instruction.

THE NATIONAL INTELLIGENCE TESTS[1]

GUY M. WHIPPLE
Bureau of Tests and Measurements, University of Michigan

HISTORICAL STATEMENT

Before the war both Professor Yerkes and Professor Terman approached the General Education Board for the support of a sort of school survey which would include the measurement of the intelligence of a good-sized group of pupils. The success of the Army Alpha Intelligence Examination made it evident that the same general methods would be applicable for such an examination of intelligence and that there would almost certainly be attempts made on the part of various individuals who had had contact with the army methods to adapt these to the examination of school children. It was felt that it would be very advantageous to the whole movement of mental testing if this adaptation could be made carefully, systematically, under the auspices of some institution or organization with prestige, and by men who would make a serious and expert contribution. The General Education Board acted favorably upon these suggestions with the proviso that the National Research Council should take the responsibility for the undertaking and that a group of four or five psychologists should cooperate in working out the details. A sum of money was appropriated for the work, and Messrs. Haggerty, Terman, Thorndike, Yerkes, and the speaker were made the members of the Committee.

We met first at Washington March 28–29, 1919, again April 29 to May 2, 1919, a third time October 17–18, 1919, and at Chicago in December, 1920. A preliminary printing of trial tests was made in the spring of 1919, and the final completed scales were issued in the summer of 1920. Something like 200,000 copies have been sold to date. The authors' royalties, I may add, are turned over to the National Research Council for use in further studies of tests.

It is my object in a few minutes to say something about the aims of our committee, the methods by which its work was

[1] Paper read at the meeting of the National Association of Directors of Educational Research, at Atlantic City, March 3, 1921.

done, the criteria we adopted in selecting and arranging the scales, and the results which are being obtained.

Scope

Our aim was to produce in a single pamphlet an arrangement of tests that could be applied to any child in the elementary school who could read well enough to participate in a group examination. In practice this means from the upper half of the third through the eighth grade. An examination that serves satisfactorily in the ninth grade or above is almost certain to be too difficult in content or in time limits to suit pupils in the lower grades. The National examination has been given with some success in the lower half of the third grade and, though the distribution tends to be "bunched" toward the lower end, yet even in that grade it operates well to locate the brighter pupils and does not give over 2 percent of zero scores. (Six in 363 were reported at Jackson, Michigan.)

Range of Difficulty

In fitting the individual tests to this proposed scope of grades and ages there was, necessarily, a problem of selecting material that would afford a proper range of difficulty. The criteria provisionally adopted in this respect, which were, I believe, successfully met with practically every test, were that *each test* should be so contrived that not more than 10 percent of perfect scores should be made by the average group of eighth-grade pupils and not more than 10 percent of zero scores should be made by the average group of upper third-grade pupils. I have not had time to verify the operation of these criteria with scores from various school systems, but I am sure that they are met, or nearly met, under usual conditions of testing.

Progressive Difficulty of Items

In order to produce a single examination that could be used both in the upper third grade and in the eighth grade and that would avoid too many zero scores in the third grade and too many perfect scores in the eighth grade, it was felt by all members of our committee that the several items within each test should vary in difficulty and should be arranged approximately in order from least difficult to most difficult.

I am aware that there are theoretical arguments against measuring speed and degree of difficulty at the same time, and that the arrangement just mentioned may be claimed to be an attempt of that sort.[2] Nevertheless, the experience of the Alpha Army Test lends support to our position; and it may be argued that if the difficulty of the items is properly distributed and the time-limit is properly set, the arrangement we have made is defensible. In any event, while we cannot argue that the steps of our scale are everywhere equal steps (while, for instance, a score of 120 may not be just 20 percent better than a score of 100), nevertheless, we can argue that the higher the score, the better the intelligence of the examinee, so that, for children of a given grade or of a given age, the total scores in either Scale A or Scale B can be regarded as indicating relative orders of merit and hence can be taken directly or indirectly as bases for classification, which is, after all, the final test of usefulness.

It remains to be said on this point that the actual arrangement of the items in the National Intelligence Tests is an empirical one; it corresponds to the order indicated by actual test results.

Coaching and its Prevention

At the first meeting of our committee there was much discussion of the danger of coaching and of the means of meeting it. Professor Thorndike, in particular, urged consideration of this point. He felt that not only would intelligence testing become a common feature of public school administration, but that it would become also a common feature of business administration, e. g., that business men would use mental tests in selecting young boys and girls for beginners in their establishments. He argued also that within some five years practically every city of over 25,000 population would establish special classes for gifted pupils and that many parents would seek to coach their children to pass intelligence examinations given for the selection of pupils for these classes. On this account, he urged that any intelligence examination that came into general usage ought to be capable of almost unlimited expansion, that it was desirable to include in our examinations only tests the material of which could be so

[2] The recently issued Ayres reading test, for instance, is based on the principle that in a rate test all items should be of equal difficulty and rate should be measured directly by quantity done within a given time limit.

extensive that it would be possible to produce even as many as thirty or forty different forms of the test just by drawing material by chance from the general reservoir of items prearranged for each test—this, even though the actual forms might not be as good as could have been produced by selecting the best of the items and confining attention to the production of, say, four or five of the best possible forms. He further argued that another way to discourage coaching would be to dump a considerable number of forms upon the market at the outset—enough to discourage any one who tried to anticipate and prepare for the examination.

The other members of our committee felt that the danger from coaching was much less serious than this proposal would imply, and there was argument for restriction of our efforts to a few forms. The compromise finally effected seems to me to meet the situation adequately. To begin with, the National Intelligence Tests appear in two independent scales, A and B, either one of which will serve well for an intelligence examination. Second, there is to be published very soon a second form of each scale (the proof is even now in hand), and this will be followed at intervals by three more forms, until five forms of each scale are available. These five forms, as will be shown later, have actually been tried with school children under proper conditions to cancel out errors of time-order. Then the forms have been edited, very slightly, altering an item here and an item there, to bring them to what is probably as near an exact equivalence as can be produced in instruments of measurement of this sort. I shall return to this point in just a moment. Third, there has been deposited at the offices of the National Research Council at Washington, complete material for five additional forms (that is, 10 in all) for each test in each scale. The material for Forms 6 to 10 was drawn by lot from material gathered by the several experimenters, just as was the material for Forms 1 to 5; there is every reason to expect that without any further trial it will afford a set of tests as nearly like the first five forms as these are like each other. Fourth, reports have been filed at Washington showing in detail the methods used by each member of the committee in preparing the items of the tests for which he was responsible. This will make it possible, if the need ever arises, to carry these methods forward and produce still other forms.

As the matter lies, then, our committee has already prepared and tested in actual use five complete forms of each scale, so that if we suppose that our five experimentally equalized forms of each scale are in print and that a prospective examinee should really sit down to prepare himself in advance to cope with any form that might be put before him, he would have to drill himself on the following:

1. 80 examples in arithmetic
2. 100 sentence completions
3. 120 logical selections
4. 200 synonym-antonyms
5. 45 symbol-digit associations (which conflict with one another)
6. 110 arithmetical computations
7. 200 pieces of information
8. 200 vocabulary items
9. 160 analogies
10. 250 same-or-different comparisons

If he succeeds in preparing these 1,865 items, "I'll say" for one that he has had a liberal education and deserves to "pass" the test.

Incidentally, it may be noted that the completion of the 10 forms necessitated the assembling, testing, and arranging in proper form and order of difficulty of 3,730 individual items, which may explain in some measure why five grown men used up nearly twenty-five thousand dollars worth of time and supplies in producing these scales—about six or seven dollars per item—a matter that always seems to puzzle the amateur mental tester and the layman who thinks that a mental test is something that can be thrown together overnight by any one with a high-school education.

Examiners

The aim of the committee as originally defined was to produce an intelligence examination, to quote the resolution adopted, "that can be used by intelligent normal school and college graduates with a reasonable amount of special training." As a matter of fact, the effort was made, however, to render the administration of the examination even easier than this statement would imply. And experience has shown that almost all classroom teachers in the elementary school can be brought, within a brief time, to administer the examination in proper style. Of

course, mistakes occur, and in any large school system there is usually one room from which the results obtained point to some error in applying the scales. This is inevitable when any sort of standardized procedure is attempted and need not be charged against the National Intelligence Tests as a fault of design. Mistakes of this sort are usually self-revealing and can be corrected by a second trial with another form of the same scale.

In Michigan cities where wholesale intelligence examining has been attempted under the direction of a central administrative officer, it has been found feasible to utilize the classroom teachers as examiners by distributing copies of the Manual of Directions in advance, with assignments of reading covering the administration of the scale to be used, then bringing the teachers together for consultation, giving them the test (with shortened time-limits), soliciting questions about details, and emphasizing certain points (like manner of giving the directions, strict adherence to the directions, careful timing, etc.) that are most likely to be overlooked or misunderstood.

Total Duration and Timing

A group intelligence examination for use in the elementary schools ought, if possible, to be capable of administration within a single period of forty minutes. Either Scale A or Scale B of the National examination can be concluded within that limit.

In the distribution of time within that limit, our committee proceeded on the assumption that a variety of comparatively short tests was psychologically preferable to a limited number, say two or three tests, of longer duration. For this reason it was agreed that no one test should exceed five minutes and that the time should, as a rule, be placed at two or three minutes without ever using fractions of a minute (this, of course, to lessen errors in timing). The final time limits are, for the five tests in Scale A: 5, 4, 3, 2 and 3 minutes, and for the five tests in Scale B: 4, 4, 3, 3 and 2 minutes, respectively.

Responses

Another principle which is essential to the success of group tests of intelligence, at least when working with pupils in the elementary school, is that responses should entail a minimum of writing, should be readily understood by the examinee, and

should be unambiguous to the scorer. Some of the group tests that have recently flooded the market take no account of these principles, which have been found in practice worthy of heed and which have contributed definitely to the usefulness of the National Intelligence Tests.

Scoring

To make a group test practically usable, the scoring must be rapid, objective, unambiguous, and simple enough to be undertaken by any clerk of reasonable intelligence. Here, again, I believe that the National examination, with the possible exception of the completion test (I say "possible" exception because even here we supply a fairly complete key) is superior to many other group tests of intelligence. Experience shows that no test can be scored without error, when the task is to handle papers in a wholesale fashion. The commonest mistakes made by unskilled scorers are (1) subtracting wrongs from attempts instead of wrongs from rights in the R − W form of scoring; (2) giving one-half as many or twice too many credits in logical selection; (3) failure to multiply loaded scores; (4) adding wrongly in computing the final score; (5) mistakes in counting up scores on a single test. The experience of Michigan users suggests that, when used in quantities, the tests should first be scored by the teachers, and that all papers should then be shipped to the office of the chief examiner and checked or rescored by clerks trained for this work. If the first half of the papers returned by a given teacher are found to be without error, the remainder may be regarded as sufficiently certain to be without systematic errors, and the work of the clerk may be limited to checking the copying of the figures and the addition for the final score.

The National Intelligence Tests are open to a small element of criticism in two points. First, the symbol-digit test (Scale A, Test 5) necessitates multiplying by 0.3. Inaccurate teachers are apt to multiply by 3, but this mistake is readily detected on rescoring. Further, the decimal that results in this test, to which some object, can be avoided by taking the nearest integer for the score. Second, the method prescribed for scoring the logical selection test when three responses were marked of which two were correct proved bothersome to many teachers. This particular

rule was formulated, I may say, over my objection (I had charge of the test in question).[3]

The keys that are provided for use in scoring have, so far as I know, met with universal approbation and contributed materially to the successful use of the examination.

Instructions

Unusual pains were taken by our committee with the preparation of the instructions for each test. Special endeavor was made to obtain simplicity, brevity, concreteness, and clarity, with emphasis upon clarity.

When clarity could not be obtained with brevity, there was no hesitation in sacrificing brevity. Thus, the instructions for the logical selection test are somewhat lengthy, and those for the symbol-digit test still more so, but this length was decided upon by actual trials that showed the necessity for every word used.

A good example of concreteness is afforded by the directions for the analogies test, which differ, for instance, from those commonly given to adults in that all reference to the term "relation" or the term "proportion" is avoided and dependence is laid upon concrete directions for conveying the idea of the task.

It is my belief that the National Intelligence Tests can claim unusual merit in the formulation of the test instructions, and that the success attained in this regard is another good example of the necessity, in the preparation of tests for children, of using nothing that will not stand the acid test of experience under working conditions.

The essence of the instructions for each test is not only definitely brought out in a fore-exercise; it is also printed concisely at the head of the test proper, where it can be consulted if need be.

Fore-Exercises

Our committee decided at the outset that each test should be preceded by a suitable fore-exercise that should consume not to

[3] I have recently checked over enough papers to prove that the simpler method of scoring that I urged at the outset is perfectly satisfactory and will not appreciably alter the scores or standards for the entire examination, except to reduce two or three points the average scores for pupils in the third and fourth grade. The forthcoming supplement to the Manual of Directions for these tests will remove the unnecessarily complicated method of scoring that was originally prescribed for this particular test.

exceed one minute and in most cases not to exceed a half minute that should be so designed as to make still clearer the nature of the task to be performed, and that should at the same time serve to equalize the knowledge with which the examinees undertake the test in case some of them may have had previous acquaintance with the test and others not.

This feature of the National Intelligence Tests has met with universal approbation and is certainly one of their points of merit.

Another feature of the instructions of the tests is the introduction, in the fore-exercise and at the head of the tests themselves, of samples in which the task is set forth and the method of responding is made clear. This feature is, of course, not peculiar to these tests. Its usefulness is apparent without argument.

Choice of Tests

At the first meeting consideration was given to practically all varieties of tests that had been proposed for use in group examinations. The experience with the army tests was canvassed and each member of the committee contributed suggestions at length. Various tests were rejected by unanimous consent, e. g., memory for digits, the maze test of Beta 1, the counting cubes test of Beta 2, etc.

Agreement was finally reached upon a list of twenty-two tests that seemed worthy of preliminary trial. These tests were the following:

Printed directions
Disarranged sentences
Arithmetical problems
Information
Opposites
Practical judgment
Number series
Analogies
Series completion
Symbol-digit
Comparison
Picture completion
Geometrical construction, or form combinations
Copying designs
Vocabulary
Picture sequence
Pictorial analogies
Recognitive pictorial memory
Sentence completion
Pictorial similarities
Computation
Logical selection

There was discussion, also, of the possibility of developing an "omnibus" test and of a test that would entail ability to organize, to carry through material demanding a wider scope of attention than necessitated by the "response" type of test illustrated in the twenty-two tests just enumerated.

The Preliminary Series

These tests[4] were prepared by various members of the committee and finally put forth in printed form in eight booklets.

Verbal A comprised six tests (arithmetical problems, directions, information, opposites, practical judgment, and analogies)

Verbal B comprised five tests (computation, vocabulary, sentence completion, disarranged sentences, and logical selection)

Non-Verbal A comprised five tests (picture completion, series completion, number comparison, symbol-digit, and form combinations)

Non-Verbal B comprised five tests (picture absurdities, copying designs, picture sequence, recognitive memory, and pictorial similarities).

The fore-exercises for each booklet were in these trials assembled together in four additional booklets—a plan which was later abandoned in favor of the insertion of each fore-exercise just in front of its test.

These four booklets of twenty-one tests were applied to pupils in the public schools of Alexandria, Virginia, Richmond, Virginia, in the Horace Mann School, New York City, and in three private schools in Cleveland, Ohio. The data from these pupils were forwarded to Dr. Truman L. Kelley, at Columbia University, along with teachers' estimates of intelligence, information concerning the age and grade standing of each pupil. Dr. Kelley, assisted by a staff of clerical workers, subjected the data to very elaborate study. In the fall his report was used by the committee to supplement its data gained in other ways concerning the merits of the several tests.

The result, without going into details, was a decision to utilize ten of the twenty-one tests that had been tried and to issue these ten tests in two batteries of five tests each, either battery as good as the other and each composed of an entirely independent array of tests.

The merit of this decision seems to me unquestioned. With five forms available soon in either scale (and ten if need be), the examiner can repeat a given battery with another form and secure virtual identity of tests and equivalence of items and score or

[4] Number series, though in the list of twenty-two tests, was not tried in these pamphlets.

he can repeat with the other battery and thus get a measurement of the intelligence of his group from a different, but equally valid angle of fire.

Non-Verbal Tests

In the preliminary series of tests a definite effort was made, as I have indicated, to devise non-verbal tests that would show satisfactory validity as measures of general intelligence—this because the committee felt, as most experimenters have felt, that there is danger of stressing linguistic ability to the point of identifying it with general intelligence.

The outcome of this attempt was disappointing. Probably some of the non-verbal tests that we tried could be further improved,[5] but we could not find more than two non-verbal tests that seemed to justify their retention, and these two have been used, one in Scale A and one in Scale B.

Power Tests

The term "power tests" has been applied, perhaps somewhat loosely, to tests in which there is no time-limit or in which the time-limit is so extended that most examinees go as far through the series of progressively more difficult items as their ability permits before the signal to stop is given. Many persons feel that there are certain forms, at least, of test in which speed should be subordinated to "power," in the sense just described. In the National Intelligence Tests, Nos. 2 and 3 in both Scale A and Scale B are planned to operate as power tests. Examiners should not be surprised, therefore, as some have been, to find children finishing their work with these tests before time is called.

Equalization of the Five Forms

I spoke a moment ago about the trial of the five forms of each scale and the minor adjustments that have been recently made to accomplish their practically complete equalization. You may be interested to know how these forms came out in this trial.

These forms were applied to six hundred pupils in grades high III to high VIII, inclusive, in certain schools at Washington, D. C. Without stopping for the scores in the individual tests,

[5] I still think, for instance, that the form combinations test upon which I experimented, has possibilities that a better form of presentation might bring out.

the total scores for the several forms before any attempt at adjustment was made, were as follows:

	Form				
	1	2	3	4	5
Scale A.....	114.9	111.8	112.3	114.4	113.6
Scale B.....	108.4	109.1	112.0	108.6	109.5

These figures furnish excellent evidence of the reliability of the method by which the several forms were prepared. The minor adjustments of which I spoke may be assumed to bring the forms to a degree of equivalence such that the total score of any form will not, at the outside, differ from the total score of any other form of the same scale by more than 0.5 points, or about 0.5 percent.

Results Obtained

The norms in the Manual.—In the first edition of the Manual of Directions tentative norms were supplied, based upon results from 2,000 pupils in Washington, D. C. and 2,000 pupils in Pittsburgh, Pennsylvania. These norms apply to pupils tested about June 1 and tested with Scale B on the day following the test with Scale A. These norms were confessedly tentative.

The results we have obtained in various Michigan cities run almost uniformly below those reported in the Manual. On the other hand, norms recently reported from East Orange, New Jersey, and those reported from certain California cities run higher than those reported in the Manual.

In comparing these results several things have become obvious:

1. Communities differ decidedly in the general level of intelligence as revealed by the tests. The results we have on file at the Bureau of Tests and Measurements at Ann Arbor leave no doubt upon this point. The figures for various Michigan cities, for instance, differ consistently from one another, grade compared with grade and age compared with age. We have concluded that each city must work out its own standards, and that probably it will be worth while eventually to work out certain state standards.

2. The norms reported from various communities are definitely affected by the composition of the groups under test. Thus,

some communities do, some do not, include pupils in speed classes. Some include negro pupils; some exclude them. Some begin at the lower third grade; others at the fourth grade. Some go only through the sixth; others through the eighth. Some have gone above these limiting grades to piece out their age standards by including pupils advanced in the grades; others have not. It follows that age standards (and grade standards, too) will have significance for comparative purposes only when the limitations of the groups under test are definitely stated and definitely allowed for.

3. The time of year at which the examination is held is a more important factor than seemed at first probable. Two graduate students in the mathematics department at Michigan volunteered to derive a formula for correcting scores to a uniform date. This work has been done, and we shall include the corrective formula in the next edition or in a supplementary edition of the Manual of Directions. Incidentally, it may be noted that these corrections are complicated by the varying ages of the groups that enter the schools in September and in February and by the proportion of repeaters in various grade groups. The correction for Scale A is 1.36 points per month. In addition, the correction for a class entering in September when compared with one entering in February is plus or minus 2.0 points, depending on which way the comparison is to be made.

Deriving mental ages and percentiles.—In illustration of the point just made—that city school systems will need to derive their own standards—I may cite the method adopted for this purpose in Jackson, Michigan.

The National Intelligence Tests, Scale A, were applied in that city to about twenty-five hundred pupils in grades III to VI, inclusive. Pupils aged ten and eleven years in the seventh grade were also assembled and tested to perfect the distributions for those ages.

Since most of the information concerning the location of children in the grades is familiar to teachers and supervisors in terms of mental age, it was felt worth while to translate the scores of the National tests into "Jackson mental ages." This was accomplished by regarding the median score of pupils of each age group as the standard score for the mental age as well as the chronological age of the group in question. Thus, all pupils aged

eight (over eighth birthday and under ninth birthday) were distributed in such a way as to locate the median and all the other deciles, and this median was regarded as indicating a mental age of 8½ years. The medians for 9½, 10½, and 11½ years were located similarly and points midway between these medians were taken as the scores indicative of mental ages of exactly 9, 10, and 11 years. The amount of overlapping was shown graphically by the percentile chart, and this chart became directly useful in locating pupils of any desired degree of deviation from the standard adopted for a given grade or group. Thus, pupils were drawn off for consideration in connection with special classes and speed classes, for double promotions, etc.[6] The chart of percentiles shown as Figure 1 is contributed by Miss Helen Davis, one of the recently elected members of this association.

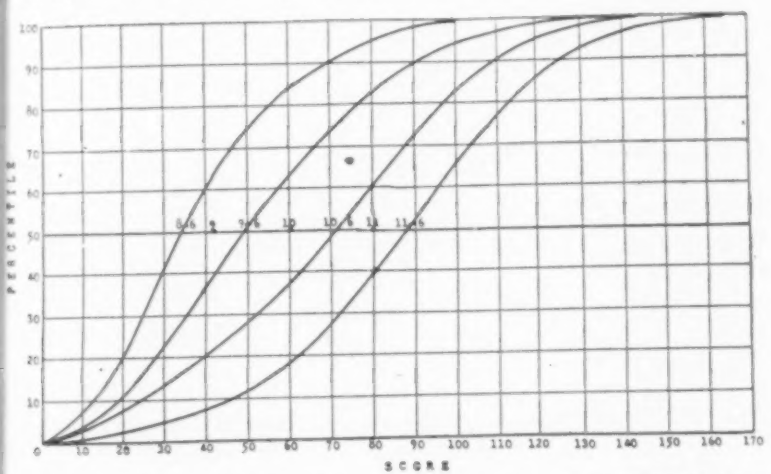

FIGURE 1. PERCENTILE CHART FOR THE NATIONAL INTELLIGENCE TEST, SCALE A, FORM I. (DATA FROM JACKSON, MICHIGAN)

Relation to Binet results.—Another possible method of determining the mental age equivalents of the National examination

[6] It will be understood that in this chart each of the four age-groups of pupils has been reduced to a theoretical 100 pupils. The figures on the base line are the scores obtained; the figures on the vertical lines are the numbers of pupils in order of excellence. Thus, in the group aged 9 years (median age approximately 9 years, 6 months) the twentieth pupil in a hundred counting from the poorest pupil scores 28, the fiftieth (or median) pupil scores 49, the eightieth pupils scores 87, etc. Or, again, 25 percent of the 8:6 group score as high as the median of the 9:6 group, etc.

scores is, of course, to equate these scores with the results of Binet examinations. I am not sure that this method is theoretically as justifiable as the one just described, namely, that of working directly to the mental age by reference to the median score. In any event, I do not have available at this moment statistics to indicate the equivalence to mental ages in the two systems of testing.

I have some figures of interest that refer to a group of thirty-two unusual cases from the Jackson, Michigan, groups just mentioned. Twenty-four of these are pupils of inferior intelligence, under consideration for transfer to an ungraded class or for withholding of promotion. Their I. Q.'s range from 58 to 88, as tested by the Stanford Binet. Their mental ages, as located by the method of determining mental ages for the National tests previously described are in 12 cases from 1 to 18 months lower, and in 12 cases from 3 to 22 months higher than the Binet mental ages: on the average, the National mental ages are 15 days lower, so that the net agreement is striking.

Five of the cases are pupils of superior intelligence, under consideration for transfer to a speed room or for extra promotion. Their I. Q.'s by the Stanford Binet range from 112 to 161. Their mental ages by the National tests are 5 and 28 months lower in two cases, and 5, 7, and 8 months higher in the other three cases. Here the National tests give a lower rating by several months on the average. In view of the fact that these mental ages are mental ages as determined by the local results in the city of Jackson, where the general scores by ages run lower than those reported in some cities, I am at a loss to account for the seeming discrepancy, though the instances are too few to be significant.

Finally, there were in this group three special cases of interest. The first was a speech defective whose Binet I. Q. was 79, but whose National Intelligence rating ran 24 months lower than the Binet mental age. The second was a pupil with an I. Q. of 90, described as a very poor reader, whose National mental age was 13 months lower than that assigned by the Binet test. The third case was a pupil whose Binet I. Q. was 112, but who is likewise characterized as unusually poor in reading. Here, again, the National mental age is strikingly discordant, being 38 months lower than that assigned by Binet testing.[7]

[7] The Binet mental age was 10 years, 8 months. The National score was only 21, or roughly that of pupils about 7½ years old.

These three cases strongly suggest, what would be *a priori* intelligible, that pupils who have special difficulty in reading will suffer a decided reduction in mental age rating by a group intelligence test when compared with their rating by the individual and oral examination of the Binet, though it remains possible that the former rating may be the more significant in predicting school progress.

They suggest, also, that pupils whose scores do not accord with what is anticipated by teachers or who are known to be possessed of some special ability or disability ought probably to be given individual examinations before any important alteration is made in their school status.

Conclusion

Speaking as a member of the Committee who devised the National Intelligence Tests, I presume I am influenced by a certain amount of unconscious bias in their favor. I have, however, sought to speak objectively, to present as directly as I could, some of the facts in the brief history of these tests, the aims of their makers, the criteria that were observed, and the results that are being obtained. If I may be permitted, therefore, to let my sentiments dictate a final sentence, I should like to say that I feel that the committee has a right to feel a tinge of pride in what it has accomplished.

THE MEASUREMENT OF LANGUAGE: WHAT IS MEASURED AND ITS SIGNIFICANCE[1]

ERNEST J. ASHBAUGH
Director Bureau of Educational Service
Extension Division, State University of Iowa

If we include under language all forms of verbal expression, we have to consider not only the most extensive of all the possible fields of pedagogical measurement but also the most complex. We have now progressed far enough in the measurement movement to recognize the fact that a single measure of any complex is quite as likely to conceal as to reveal, and that it is of very little value from the standpoint of diagnosis and remedial work.

This paper confines itself to the field of written expression. It is realized that many other phases might well be discussed under the term "measurement of language," but a desire to keep the paper within the probable time allotment, and to speak very definitely rather than in a general way, forces this limitation.

The various scales for the measurement of composition illustrate very well the above statements. The Hillegas Composition Scale, the Thorndike Extension of the Hillegas Scale, and the Nassau County Supplement by Trabue were all designed to measure general merit with the consequent result that the scales are very difficult to use and the result obtained does not greatly assist the child or the teacher in remedying defects and in thus progressing toward the goal desired. It is to be remembered that the case under discussion is the measurement of language, and what is said here and will be said later pertains directly to the use of these scales as measuring instruments.

Ballou in developing the Harvard-Newton Scale attempted at least partially to remedy this defect by constructing separate scales for each of the four forms of composition. However, any one who has read many children's themes will instantly recall that children seldom write themes which are wholly narrative, descriptive, argumentative, or expository. They have a very disconcerting way of mixing two or more forms into one inglorious whole.

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

The latter scale does attempt to be of assistance to the teacher or supervisor in one way in that certain merits and demerits of each of the samples in the scale are pointed out. This is certainly a commendatory feature, but the scoring of compositions for general merit even with these aids is still far from an objective matter. Only careful intensive training will develop one to the point where he can be fairly sure his scoring will have slight deviation from the norm or established value.

The composition scale developed by Breed and Frostic for the sixth-grade attempts a narrower range of merit with consequently, a greater homogeneity, but in other respects it adds little or nothing to the solution of the problem beyond that accomplished by the scales previously named.

Willing made a distinct contribution when he provided in his scale for separating the mechanical elements of capitalization, punctuation, spelling, and paragraphing from the thought element. After having done this, however, he seemed to have felt under the necessity of combining for a single score; and thus, so far as the score itself is concerned, it is no more enlightening than it would have been if he had not separated the two factors in the first place. The accuracy with which the score is secured is probably greater; but whether the compostion is poor because of mechanical defects or because of weakness in thought elements is completely hidden.

What has just been said concerning the scales for measuring composition is not to be construed as an argument against their use. A trained person using a scale can doubtless obtain valuable information concerning the general merit of compositions written by a given class; and this information may be rendered still more valuable when compared with the norm for children of the same grade elsewhere. There is some evidence that scales designed to measure general merit may be used rather effectively as a teaching device in much the same way that the various handwriting scales have proved to be valuable. When samples from the scale are placed upon the wall of the room and the children are encouraged to compare their own products with them, they seem to be able to study these samples and modify the quality of the compositions in terms of the general qualities considered in the scale.

What has been said is only designed to point out the nature of the product which is to be measured and the fact that its com-

plexity makes the measuring instrument of very little value fr the standpoint of diagnosis and remedial work.

Any analysis of written expression (which of necessity must the phase of language ability which will be most frequently me ured) must recognize at least three groups of factors, nam mechanical, grammatical, and rhetorical. Each of these gro contains many factors or elements, and only as we separate th complexes into simpler and simpler elements will our measureme become truly helpful to the teacher and supervisor in the impro ment of the work in the classroom.

Trabue in his Completion Test Language Scales attempted devise an instrument "for the measurement of ability along c tain lines closely related to language." He states that "Ebbi haus who invented the completion test method characterized as 'a real test of intelligence.'" Those who have used these sca have been more inclined to characterize them as intelligence te than as tests of language ability. Using Scale A Trabue repor a correlation with Hillegas Composition of 0.72 for a group of seventh-grade children, a correlation of 0.85 with the Binet te for a group of 39 boys, and a correlation of 0.74 with the Bin tests for a group of 50 boys and girls. Dr. H. A. Greene of t University of Iowa, using Trabue's Scale B and Hillegas comp sition scores from 132 high-school students in one school, found correlation of only 0.38. In another high school the same te gave a correlation of 0.14 for a group of 58 students.

From these data it would seem that native ability rather th school training is the determining factor in accomplishment this test. Possibly the same is true, though to a lesser degree, achievement in writing compositions. To the extent that scal measure intelligence rather than training, their use in the solutio of the teacher's problems in language instruction must be negati and general rather than positive and specific.

Greene's Organization Tests are the result of a definite attem to analyze language ability and to present an instrument for th measurement of a single phase of this very complex problem. H does not claim to have reduced the problem to the measureme of a simple element. The ability to take a limited number of ide presented in jumbled order and to rearrange them in logical ord is evidently closely related to language ability and is certainl simpler than the ability involved in writing a composition.

He found a correlation of 0.52 with composition for a group of 135 high-school pupils and a correlation of 0.47 with composition for a group of 109 elementary-school pupils. Since his correlations with Terman's mental age (individual measurements) were 0.41 and 0.50 respectively with high- and elementary-school pupils, he concludes that his test is a better measure of the elements involved in language ability than is Trabue's Completion Test. In a recent article in the JOURNAL OF EDUCATIONAL RESEARCH Brooks stated that this test is "mostly an intelligence test" but failed to present any evidence in support of his statement.

Granted that this test will measure somewhat accurately the phase of language ability which it is intended to measure, the fact remains (with this test as with many others) that teacher and supervisor remain quite as much in the dark after as before its use concerning what should be done to improve this ability. A knowledge of low scores is not especially helpful unless at the same time remedial treatment is at least suggested; and thus far, practically nothing is known concerning the type of teaching or drill work which will improve language ability and result in greater achievement on a later test.

Starch, in his Grammatical Scale A, limited himself to the factor of grammatical errors. The scale presents two forms of expression, one correct and the other incorrect, and the child is to discriminate between them, crossing out the incorrect form. The task is thus very definite, the distinguishing between correct and incorrect grammatical forms. The sentences are grouped in fours. These groups are called steps, and the successive steps are supposedly of equal increments of difficulty. The scoring is in terms of these steps of unit value.

Rather extensive use of the scale has brought out the following facts:

1. The test shows rather definitely the sentences in which the pupil cannot distinguish between correct and incorrect form and thus furnishes the teacher with a clue to the instruction needed.

2. The ability to recognize correct form seems to be more closely related to language habits than to knowledge of technical grammar.

3. Intergrade differences are so small when using Starch's method of scoring that they become insignificant and hence the score means little.

4. On the basis of returns from children in grades seven and eight, there seem to be rather wide differences in the difficulty of individual sentences within the same group. (See Table I. Hence, some other scheme of scoring would probably be of greater value. The facts given in Table I represent a larger number of

TABLE I. STARCH'S GRAMMATICAL SCALE

Step	Sentence	Accuracies (Percent)	
		Grade VII	Grade VIII
5	1	85.0	94.0
5	2	93.0	96.0
5	3	78.0	75.5
9	1	56.5	74.0
9	2	74.0	75.0
9	3	44.0	54.5
9	4	81.5	84.0

well distributed cases than Dr. Starch used in his formation of the scale. The sentences from only two steps are given but several others show similar conditions, namely, that one or two of the sentences are much more difficult than the others within the same step.

5. The scale covers a number of types of grammatical errors but the author has furnished no data showing the basis of selection of the types and the relative importance of each.

Charters' Diagnostic Language Tests, one each for pronouns, verbs, and miscellaneous, attack the same problem attempted by Starch, but with a few noteworthy variations.

1. The material was secured from an extensive survey of language usage among children and was grouped on the basis of the three classifications noted above.

2. Instead of the place of error being pointed out and a choice made necessary between a correct and an incorrect form, the child is confronted with a series of sentences each of which he is told to mark with a cross in case he thinks it correct and to mark out the incorrect form and write the correct one in case a correction is to be made.

3. The test is scored in terms of single unit value for each sentence marked correctly.

4. The test for pronouns purports to present each of the different pronominal uses, and thus covers completely this phase of language ability. The following illustrations show how the test will point out clearly to the teachers the errors which the children fail to recognize as incorrect. The following sentences were pronounced correct by more than 75 percent of a large group selected at random from eighth-grade classes. The sentences are numbered as they appear on Dr. Charters' Pronoun Test.

3. It teaches a person something you may use.
5. When one lives in town they hear noises.
13. Are those them?
28. That's her.

The following sentences were pronounced correct by more than 50 percent of the same group of eighth-grade pupils.

7. Who do you want?
14. It was only us.
21. He pushed John and I.
25. Annie called to you and I.
27. Was that him?

The use of the test with grammar-grade pupils shows clearly that a number of these type errors persist with a very large proportion of the children who are practically ready to leave the grades, and the individual test papers will point out definitely to the teacher the work which needs additional attention. It is probable that the sentences should be given a weighted score on the basis of a combination of frequency of social usage and the inaccuracy of eighth-grade pupils, in order that their relative importance may be more forcibly impressed upon those who take the test.

A second edition of Charters' tests provides additional spaces beside the sentences. In this space the child is to write the grammatical reason for the changes which he makes in correcting the sentence. This affords a measure of the child's ability to give the correct grammatical rule or principle which applies to the case in point. He must not only know the principle but also be able to recall it in its proper relationship.

Kirby has attacked the same problem as Starch and Charters. He presents correct and incorrect forms for the child to choose between in the same manner that Starch does. Kirby, however, scores on the basis of unit value for each sentence. No evaluation

has been made for either difficulty or social usage. Here, as was stated concerning both Starch's scale and Charters' test, the children tend to reveal quite accurately the types for which they do not distinguish correct from incorrect language forms. In this test, however, no provision has been made to care for the operation of chance.

In addition to the check on the language form, Kirby attempts to measure the knowledge of grammatical principles involved in a different manner from that used by Charters. Beside a set of sentences there are given the different grammatical principles involved in this set of sentences. These, however, are arranged in such a way that a principle is never directly beside the sentence whose usage it governs. The child is therefore called upon not to recall from his own memory the grammatical principle which will justify the correction he has made, but to select from a number of grammatical principles which are before him the one which will apply in the particular case.

As was shown in the case of the Charters' tests some of the sentences involved language usage which is incorrectly checked in a large proportion of cases, while others are seldom missed. For example, it was found with two hundred seventh- and eighth-grade children that the sentence "It is (I) (me)" was missed but by 2 percent of the children; "(Mr. Smith) (Mr. Smith he) went ahead" by only 1 percent; "There (was) (were) many reasons for his actions" and "He (don't) (doesn't) belong in that group" by but 7 percent in each case. On the other hand, "He is the man (who) (whom) you said was injured," "I (lay) (laid) on the sand two hours yesterday," "It is a slight to me who (has) (have) always been your friend," and "We admire (that) (those) kind of people" were missed by more than 50 percent of the children on the language side alone.

On the technical grammar side, the principles governing the following sentences—"That book is not (hern) (hers), "(May) (can) I bring the next story to read?," "(Leave) (Let) me go with you," and "He divided his money (among) (between) his four brothers"—are the only ones which were correctly checked by practically all of the children. At present it would seem that the tendency in language work in our elementary schools is away from technical grammar with corresponding emphasis upon formation of correct language habits. Wherever this attitude is believed to

be the important one, the second phase of this test may well be omitted, since any test tends to emphasize the type of work which will get the results measured by the test. If the supervisor does not wish his teachers to stress the knowledge of grammatical principles, he should not measure the children's lack of knowledge of such principles.

It is true that ability to recognize incorrect language forms and to correct them does not guarantee that the child will use the correct form in either his oral or written language. Nevertheless it is quite certain that whenever the child cannot or does not recognize incorrect forms when presented to him for discriminating effort, he is not likely to use the correct form when his focus of attention is upon the general thought as is the case in either oral or written composition.

The purpose of this paper has been accomplished if the following facts have been clearly pointed out.

First, that the trend in the measurement of language has been from a gross measure of general merit to a specific measure of individual factors which are included in this general merit.

Second, that different efforts have been made to analyze the field and to discover through a study of social usage the phases which need careful attention.

Third, that tests which bring sharply to the attention of teacher and pupils the strength and weakness of each individual in the phase of language measured are more valuable than those which fail to reveal these situations.

Fourth, that while a considerable amount of work has been done, it is evident that our instruments of measurement are not yet perfect, that the field has not been completely covered, and that, therefore, although we are on our way, there remains yet a tremendous amount to be done.

INTELLIGENCE TESTS IN CLASSIFYING CHILDREN IN THE ELEMENTARY SCHOOL

CHARLES FORDYCE
University of Nebraska

For the purpose of determining to what degree the school supervisor may through mental tests predict the educational career of public school children and thus throw light on their proper classification, the writer made a study of the results of the Haggerty Intelligence Examination in comparison with the school grades and estimates of teachers in the case of a group of pupils in the elementary grades at Lincoln, Nebraska. The tests were given to 1,078 grade children distributed as follows: primary grades (I, II, and III) 602 pupils, intermediate grade (VI) 476 pupils. The Haggerty Intelligence Examination, Delta I was used in the primary grades and Delta II in the intermediate grade. The tests were given by Miss Clara Slade, psychologist in the Lincoln schools, and the results were tabulated and interpreted by a class of 67 graduate students in educational measurements under the direction of the writer. The test paper of each pupil was corrected by three different students working independently. The examination was given in the middle of May, 1920.

All teachers having to do with the instruction of the children gave their ratings of the intelligence of the pupils at the time the tests were made. The judgment of the teachers was based upon class grades made during the year and upon their general estimate of intelligence. In the case of each pupil the teacher's rating represents the average of the combined estimates of several instructors. As will be seen in Table I, the intelligence rank, as given by the teacher, was arranged on a five-point scale: very superior pupils being ranked A, superiors B, average pupils C, inferiors D, and very inferiors E. The rating by the intelligence tests was arranged on a similar scale and was based on the I. Q.'s obtained in the usual way (that is by dividing the mental age by the chronological age). Rank 1 in intelligence represents the students whose I. Q. was 120 or above, rank 2 those with an I. Q. from 110 to 119, rank 3 those with an I. Q. from 90 to 109, rank 4 those with an I. Q. from 80 to 89, and rank 5 those with an I. Q. below 80.

In Table I all entries in the upper left-hand corners of the compartments are for the sixth-grade pupils; those in the lower

TABLE I. CORRELATION BETWEEN TEACHERS' ESTIMATES AND INTELLIGENCE QUOTIENTS

I. Q's	Teachers' Estimates E	D	C	B	A	Totals
80	9 32 23	49 84 35	28 50 22	4 9 5	0 0 0	90 175 85
90	3 11 8	26 79 53	42 72 30	12 21 9	1 2 1	84 185 101
110	2 15 13	32 118 86	139 263 124	52 132 80	9 13 4	234 541 307
120	0 1 1	0 8 8	14 26 12	16 47 31	12 19 7	42 101 59
	0 1 1	1 5 4	5 22 17	10 24 14	10 24 14	26 76 50
Totals	14 60 46	108 294 186	228 433 205	94 233 139	32 58 26	476 1078 602

Correlation, +0.44; P. E., 0.02
Figures in upper left-hand corners of squares are for sixth-grade pupils, total 476.
Figures in lower right-hand corners of squares are for pupils of grades I, II, and III-A, total 602.
Figures in center of squares are for all pupils, total 1,078.

right-hand corners are for the primary-grade children; while those in the center of the compartments are for all children—i.e. the entry in the center of each compartment is the sum of the other two entries in the same compartment. The detailed tabulation given in Table I shows that for 445 of the pupils or 41 percent, the ratings of the teachers and those of the intelligence examination were the same. According to the intelligence rating the teacher overestimated 383 cases of those of average or inferior capacity. This tendency is found to be common and due mainly without doubt to the fact that many of the average or inferior pupils are over aged. On this account, while they are inferior in mentality and may be doing only average work in the grade in which they are placed, yet they have the maturity of body and the instincts and emotions of children of their own age. They are therefore

estimated to a considerable degree by these characteristics. The overestimation is also probably due in part to the fact that inferior pupils are in many instances attractive, vivacious, and talkative, always ready to respond to the teacher. These qualities usually elicit a good grade.

Of the average or superior pupils 220 were underestimated by the teachers. This may be due partly to the fact that the superior student is in many instances young for the grade in which he is placed. He is, therefore, immature in bodily development and in his instinctive and emotional reactions, and is in some degree judged by these qualities. The underestimation of capable children may also be partly due to the fact that the work is so easy for the superior boy or girl that it does not challenge his efforts. As a result of this he often neglects his assignments, and is consequently rated average or low by the teacher.

It is a significant fact that among the 1,078 pupils, 74 percent of the highest ranks given by the teachers were given to pupils who were young for their respective grades. Such pupils had higher Intelligence Quotients, and being mentally superior to their fellows should be expected to do superior work. Ten percent of the highest grades were given to the over-aged whose intelligence quotients were low. This doubtless was due to the prevalent tendency on the part of teachers to overestimate the older pupils. Of those ranked average by the teachers 60 percent were of average rank as shown by the mental test, thus corroborating the fact usually found that the average pupil is more correctly judged and ranked by the teacher than either the superior or the inferior. Of the average grades given by the teachers 28 percent were given to pupils of inferior mentality as shown by the mental test. Of the lowest grades given, 4 percent were given to superior children. This is due no doubt to the tendency to underestimate the superior or younger pupils or to the fact that these superior pupils find the work so easy that they do not put forth an effort to do it. One of the greatest values of the mental test is shown in this latter case. It discovers the pupil who should be doing superior work and leads the teacher in many instances to bring about a procedure which will stimulate such students to do better work.

The highest rankings by the teachers were made in the case of those of highest intelligence quotients; of the students with

intelligence quotients below 90, 57 percent were given low ranking by the teachers, and only 9 percent received above average ranking. One-half of one percent were placed by the teachers in the highest rank.

The median I. Q. for the entire group was 100.5. The boys ranked higher in the mental tests than the girls, but the girls were rated slightly higher by the teachers. The median I. Q. for the boys was 105, and for the girls 96, but the median rating by the teachers was 85 percent for the boys, and 86 percent for the girls. It is difficult to explain this discrepancy in the correlation. The correlation between the ratings by the mental test and by the teachers was 0.44 with a probable error of 0.02. This correlation is not as high as we secured in the case of similar tests in the eighth grade and the high school.

Conclusion

This study in the elementary grades of the Lincoln schools indicated:

1. That the boys ranked slightly higher in the mental tests than the girls, but that the girls ranked somewhat higher by the teachers' estimates. (This is contrary to the results obtained in the Teachers College High School where the girls ranked superior, both by the mental tests and the estimates of the teachers.)

2. That the teachers tended to overestimate the inferior student and to underestimate the superior.

3. That the correlation between the mental tests and the teachers' estimates was sufficiently high to indicate that it is possible for one to discover by means of mental tests along with the grades and estimates of teachers such capacities as are essential in determining the proper grading and promotion of pupils.

estimated to a considerable degree by these characteristics. The overestimation is also probably due in part to the fact that inferior pupils are in many instances attractive, vivacious, and talkative, always ready to respond to the teacher. These qualities usually elicit a good grade.

TIGHT BOUN

pupils and the work so easy that they do not put forth an effort to do it. One of the greatest values of the mental test is shown in this latter case. It discovers the pupil who should be doing superior work and leads the teacher in many instances to bring about a procedure which will stimulate such students to do better work.

The highest rankings by the teachers were made in the case of those of highest intelligence quotients; of the students with

intelligence quotients below 90, 57 percent were given low ranking by the teachers, and only 9 percent received above average ranking. One-half of one percent were placed by the teachers in the highest rank.

The median I. Q. for the entire group was 100.5. The boys

THE INTELLIGENCE EXAMINATION FOR HIGH-SCHOOL FRESHMEN[1]

IRA J. BRIGHT

Superintendent of Schools, Leavenworth, Kansas

We had four general purposes in mind in carrying forward the study of the intelligence examination for high-school freshmen. The first was to determine the extent to which an intelligence examination can be used at the time a pupil enters high school, as a means of forecasting his probable success in the various high-school subjects and in order that he may be given intelligent guidance in his program-making. The second was to provide the best possible criteria on which to base the organization of class groups. The third was to provide the instructors in the high school with a definite statement of the mental abilities of the individual members of their respective classes, in order that they might better adjust their methods of teaching, the subject matter of their courses, and their requirements to the needs of the individuals in their respective classes. The fourth purpose was to determine the adaptability of the Terman Group Test of Mental Ability to first-year high-school pupils.

I am sure no argument is necessary among school men to prove the statement that up to the present time there has been very little guidance given to high-school freshmen in their program-making. On entering the high school the child has the opportunity to elect certain courses or subjects. When advice is given by advisory committees, class sponsors, or principals, it is given in most cases without any definite knowledge on the part of the advisor of the intellectual equipment of the child.

I know certain principals who always advise those who complete the eighth grade in their buildings to take Latin and algebra in the high school regardless of the mental capacity of the pupils. A review of the failures by subjects from the Leavenworth superintendent's reports for years past shows an excessively large number of failures in both Latin and algebra. This is no doubt due to the fact that a few years ago the distinctive disciplinary value of Latin and algebra was unquestioned. It was

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

thought that these subjects were worth while for the general mental training which they provided regardless of all other considerations. The importance given them by tradition is still effective in drawing great numbers into these courses who are sure to meet with discouragement and failure because they lack the mental power to master them. That the loss in the freshmen year of high school, due to the lack of an intelligent and scientific method of guiding pupils in their program-making, is tremendous there can be no question.

The scholarship grades used in this study were those given by the high-school teachers at the end of the first quarter of this school year. The intelligence examination was not given until two weeks after the quarterly reports of the teachers were handed in and the results of the intelligence examination were not returned to the teachers until all information from the teachers prompted by this test had been secured. This precaution was taken so that there could be no conscious or unconscious prejudice influencing the teacher in the distribution of the scholarship grades or in answering questions concerning certain individuals about whom the writer sought further information.

The tests were all given by the high-school principal after he had gone over them with the writer and after he had made a careful study of the instructions. A stop watch was used so as to avoid discrepancies in the time element of the test.

The grading was all done in the superintendent's office under his immediate direction and supervision. The statistical work was done by the writer himself.

The correlation coefficients mentioned in this study have reference to the Pearson Product-Moment Coefficient of Correlation. Most of the coefficients were computed by the long method applying the formula $r = \frac{\Sigma (xy)}{\sqrt{\Sigma x^2 \, \Sigma y^2}}$ and were checked by applying the shorter method suggested by Dr. Ayres in the April number of the JOURNAL OF EDUCATIONAL RESEARCH.

In the correlation tables showing the distribution of teachers' marks and intelligence scores, six steps were used for the scholarship grades and fourteen for the intelligence scores. In the correlation tables showing the distribution of scores in each of the ten tests of the Terman examination and in the complete test,

from ten to twelve steps were used for the individual tests a fourteen for the total score.

Table I shows the distribution of the freshmen according the teachers' marks in Latin and according to intelligence scor The median intelligence score of all Latin I students was 12 while the median intelligence score for all freshmen was 97. T shows that some selective agency had been operative. The were, altogether, nine in the freshmen class whose intelligen scores were between 40 and 49 on the Terman scale. Fortunate only one of these elected Latin. There were also nine freshm in the next lowest group—i.e., those with intelligence scor between 50 and 59. None of them elected Latin. Out of the te in the next lower group only two were taking Latin. At the oth end of the distribution the result of selection is quite as notic able. In the entire freshmen class there were three whose intel gence scores were between 170 and 179. All three of them we taking Latin. There were also three in the 160 to 169 group, an two of the three were taking Latin. In substance the student with the higher intelligence scores elect Latin while those with th lower intelligence scores do not. The selection here shown is du primarily to two causes. First the general feeling among pupil that Latin is hard. The brighter pupils not having had trouble in the elementary schools in making their grades are less effected by this feeling. Those of lower intelligence have more reason t be afraid of a hard subject. The other cause is the fact that we have given some attention the past few years to the guidance o pupils in the selection of their courses. We have attempted t keep out of the Latin classes those who are rated very poor by their principals. While the attempt to guide pupils in this matter has been inadequate, it has been somewhat effective.

The distribution on the intelligence scale of those making the different grades in Latin indicates in a general way that the correlation between teachers' marks in Latin and the intelligence scores is high. The coefficient of correlation is $+0.65$.

The complete distribution shows that 86 percent of all those whose scholarship grades in Latin were above 90, had intelligence scores above the 75-percentile, that is they were in the upper quarter of the whole freshmen distribution on the Terman Group Test. All first-year Latin students whose intelligence scores were below the 25-percentile received failing grades in that subject.

TABLE I. TERMAN GROUP TEST OF MENTAL ABILITY AND TEACHERS' MARKS IN LATIN

Intelligence Score	Teachers' Marks in Latin I					Total
	70–74	75–79	80–84	85–89	90–94	
170–179				1	2	3
160–169				2		2
150–159	1		1	2	1	5
140–149			1	1		2
130–139		3	1	1	1	6
120–129	1		7	2	2	12
110–119	2				1	3
100–109		2	1	1		4
90– 99	1			1		2
80– 89	2	3	1	2		8
70– 79	2	3		1		6
60– 69	1		1			2
50– 59						
40– 49	1					1
Total	11	11	13	14	7	56
Median	87.5	86.3	125	130	135	121.7

Only 15 percent of those whose intelligence scores were below the median of the group received scholarship grades in Latin above 80. Or stating it conversely, 85 percent of the first-year Latin group whose intelligence scores were below the median of the whole freshmen group received the lowest passing grade or failed altogether.

I believe therefore it is a safe prediction to say that those whose intelligence scores in the Terman Group Test are below 76, the 25-percentile of the entire freshmen distribution, have absolutely no chance to make a passing grade in Latin; and that those whose intelligence scores are below 97, the median of the whole group, will do unsatisfactory work in Latin. And I believe the prediction will hold good for almost any high school, that all freshmen whose intelligence rating according to this test places them in the lower quarter of the entire freshmen group will most likely fail, and that those whose intelligence scores are below the median of the freshman group will most likely do unsatisfactory work.

Let us now consider the distribution of English grades and intelligence scores as shown in Table II.

TABLE II. TERMAN GROUP TEST AND TEACHERS' MARKS IN ENGLISH

Intelligence Scores	Teachers' Ratings in English						Total
	70–74	75–79	80–84	85–89	90–94	95–100	
170–179					2	2	4
160–169						2	2
150–159				1	4	2	7
140–149	1	2			3	3	9
130–139		2	3	3	5	1	14
120–129			1	6	5	2	14
110–119	2	3	1	3	1	1	11
100–109	3	2	4	3	6	1	19
90– 99	3	2	5	7	2		19
80– 89	2	5	2	3	5	2	19
70– 79	2	7	7	6	1		23
60– 69	3	2	2	1			8
50– 59	1	1	2	3	2		9
40– 49	3	1	1	1			6
Total	20	27	28	37	36	16	164
Median	85	85	90	92	122	143	99

Table II in a general way indicates that there is some positive relation between teachers' marks in English and intelligence scores. The coefficient of correlation is +0.72.

A more detailed study of the complete distribution will reveal some facts not precisely indicated in the Table II. It shows that 70 percent of all those who received the highest scholarship grades in English also had intelligence scores in the upper quarter of the entire freshmen distribution. On the basis of Table II the chance is about 1 to 25 that those whose intelligence rating places them below the upper quartile will receive the highest (95-100) scholarship grades in English. Moreover, of those receiving failing grades in English, 40 percent were in the lower quarter of the intelligence distribution and 70 percent were below the median.

These facts warrant the statement, I think, that the kind of grades those who enter the English classes of the first year of high school will likely make, can be rather well determined in advance by the Terman Test.

The correlation table (Table III) showing the distribution of algebra grades and intelligence scores tells much the same story except that the correlation is a little lower, the coefficient being +0.50.

TABLE III. TERMAN GROUP TEST AND TEACHERS' MARKS IN ALGEBRA

Intelligence Scores	Teachers' Ratings in Algebra						Total
	70–74	75–79	80–84	85–89	90–94	95–100	
170–179				1		2	3
160–169				2		1	3
150–159					1	5	6
140–149		2	2	1		3	8
130–139	1	2	3	2	4		12
120–129		1	1	2	4	5	13
110–119	3			1	2	2	8
100–109	2	3		4	4	4	17
90– 99	2	3	2	4	5		16
80– 89	4	4	5	3	1	2	19
70– 79	2	8	3	3	3	1	20
60– 69	3	4	1	1			9
50– 59		5	1	2			8
40– 49	2	1	1	1			5
Total	19	33	19	27	24	25	147
Median	86	78	87	99	107	127	98

From the more detailed distribution we find that 88 percent of those receiving the highest scholarship grades had intelligence scores above the 75-percentile of the entire intelligence distribution. Also, 71 percent of those who failed or received the lowest passing grade in algebra, had intelligence scores below the median of the whole freshmen group. Only 30 percent of those in the algebra classes whose intelligence scores were below the median of the whole freshmen group received grades in algebra above 85, that is, grades in the three upper scholarship groups.

It is obvious that a large number of our first-year high-sch children are enrolled in algebra who, on account of their low inte gence, can never receive material benefit from it. There practically no chance for those in the lower quarter of the inte gence distribution to do satisfactory work in algebra, and ve little likelihood that those whose intelligence scores are belo the median will get much value out of the subject. To my min it is positively absurd to assume that all high-school pupils shou take the course in algebra. Certainly if all are required to ta this course, the subject matter, requirements, and methods shou be adapted to the varying mental abilities represented in th freshmen group.

The distributions in the other academic subject are simila to those already discussed, and so I shall pass on to the distribu tion of grades in the handicraft subjects. I have included i this group manual training, domestic art, domestic science, fre hand drawing, and penmanship. The coefficient of correlatio between ntelligence scores and teachers' marks in handicrafts i only +0.36. This is very much lower than the correlation between the academic grades and intelligence scores. Th significance of this fact cannot be escaped. It tells us plainl what to do with those of low intelligence who enter the hig school. It is clear that this group should be encouraged to tak handicraft courses and that their enrollment in Latin and algebra especially should be discouraged.

The high mortality rate in the first year of high school is no doubt due largely to discouragement which this group is bound to receive, especially in the academic subjects, under the present methods of administering the courses. When teachers can be provided w th an accurate index of the mental abilities of their pupils and when they have the ability themselves to adapt subject matter for the benefit of their respective classes, the number of failures will be greatly reduced if not entirely eliminated. We have as far as possible organized the classes for first-year high-school pupils who enrolled the second semester on the basis of intelligence scores. The enrollment in the midyear was small but we were able, for the most part, to put those whose intelligence scores were above the median of the group into one section and those whose intelligence scores were below the median of the group into another section. • From all reports so far received the differ-

ence in the quality and quantity of work being done by the two sections is marked.

Where it is possible the classes in any subject should be organized on the basis of three standards; those in the upper quarter of the intelligence distribution comprising one division, those in the middle 50 percent of the distribution comprising another division, and those in the lower quarter, a third. The possibilities of the classes comprised of children of superior intelligence are almost unlimited. The individuals in these classes can be held responsible for a much higher quality of work as well as for a much greater quantity of work.

For the classes in the lower division the courses should be simplified and these groups held only to standards adapted to them. It is unreasonable to expect the best results in classes where there is represented such wide variation in mental abilities as is now found.

To my mind one of the greatest values which come from the use of intelligence tests is the stimulative effect of the result on the individuals tested, especially those making the higher scores. The recognition on the part of the individual that he stands high on the intellectual scale makes an appeal to his pride that is more effective than any other appeal that can be made.

Since giving the intelligence examination in our freshmen classes, the principal of the high school has called into conference those whose intelligence rating was high but whose scholarship grades were low, pointing out to them the fact that they are not in any sense measuring up to their possibilities and appealing to their sense of pride to raise the standard of their school work. This is proving wonderfully effective and will always prove effective when the principal or teacher is a person of strong personality and has the confidence of the student body.

On the other hand, I believe the greatest possible caution should be exercised in the use of the results of mental tests. In the first place, teachers must learn that the results secured from any one test may, in individual cases, be in error. No teacher should jump at the conclusion that because a child makes a low score in a test he belongs on the low intellectual plane indicated by the score. Judgment in these cases should be held in abeyance until the test result is verified or disproved by other tests and the further reactions of the individual. High scores in an intelligence

test are more likely than low scores to be true indications of mental ability. Then, too, an over-evaluation of mental ability is not nearly as serious as an under-evaluation of ability.

To get a further check on the validity of our test results we sent out to the teachers concerned a list of the names of the pupils who had high intelligence scores and low scholarship grades and of those who had low intelligence scores and high scholarship grades. The list was accompanied by a questionnaire in which the teachers were asked to indicate which of the following factors were most responsible for the scholarship grade received by the pupil: mental ability, attendance, health, indifference, laziness, interest, effort, attention, energy, affability, application. As an example let us take the report on one of these cases. Here is a boy whose intelligence score is 54 but whose scholarship grades are medium. Three of the teachers say that he hasn't much mental ability; one says he has ability. The first three teachers say that his scholarship grades are due to interest, application, and attention. In this case we have the judgment of three teachers against one that the test evaluation is about right.

To summarize the reports of the teachers, 85 percent of all those whose intelligence scores were in the upper quarter of the whole freshmen distribution but whose scholarship grades were low, were reported by the teachers in this "follow up" questionnaire as having mental ability. The low grades were accounted for in this report by lack of application, laziness, inattention, poor attendance, and the like. There were 29 in this group with high intelligence scores and low scholarship grades. Of these 25 were credited with high mentality in the "follow up" report. There was, then, a material disagreement between the test evaluation and the teachers' judgment in only four cases of the 45 in the upper quarter.

There were only seven of the 45 in the lower quarter of the intelligence distribution whose scholarship grades were above 85, that is, in the three higher scholarship groups. In the individual reports of the teachers, four of these were rated as having low intelligence. So there were only three of the 45 on which the intelligence test evaluation and the evaluation of the teachers differed.

In short, the Terman Group Test of Mental Ability has given us a distribution that is most consistent with the teachers'

judgment of mental ability as represented in the scholarship grades and as represented in the analysis of the individual cases.

It is evident, then, that the most influential factor in the determination of scholarship grades is mental ability. The greatest help that can be given a teacher is an accurate evaluation of the mental abilities of her pupils. If the classroom work of a pupil does not measure up to his mental ability, the cause should be immediately sought and eliminated. It is a tremendous loss to the individual capable of making the highest scholarship grade to be permitted to go through school with only average scholarship attainment. There is a tremendous loss to society as well as to the individual in every case where superior ability is not recognized and is not given the best possible opportunity for development.

I am thoroughly convinced that the efficiency of high-school teachers can be immeasurably increased by the judicious use of some such intelligence examination as the one used in this study. It is possible, however, that a shorter examination may be sufficient for the purpose. To determine whether or not the Terman Group Test could be shortened without impairing the results, we found the coefficient of correlation between each of the ten tests and the whole examination and between the first

TABLE IV. CORRELATION BETWEEN EACH TEST OF THE TERMAN GROUP TEST AND THE WHOLE EXAMINATION

Test	Correlation Coefficient
1	0.77
2	0.62
3	0.82
4	0.71
5	0.66
6	0.51
7	0.75
8	0.51
9	0.52
10	0.75
1–5	0.90

five combined and the whole. These correlations are shown in Table IV in the form of coefficients. Table V shows the details of the correlation between Tests 1 to 5 and the entire test.

TABLE V. TERMAN GROUP TESTS. CORRELATION BETWEEN TESTS 1–5 AND TOTAL SCORE

Scores on the Entire Test	Scores on Tests 1–5											Total
	16–23	24–31	32–39	40–47	48–55	56–63	64–71	72–79	80–87	88–95	96–103	
170–179										3	2	5
160–169									1	2		3
150–159								1	6			7
140–149							1	5	4			10
130–139						1	5	6	2			14
120–129					1	4	7	2				14
110–119				1	2	5	3					11
100–109				3	5	10	2					20
90– 99				4	9	8						21
80– 89			1	12	10							23
70– 79		4	12	8								24
60– 69		5	4	1								10
50– 59	5	4										9
40– 49	5	4										0
Total	10	17	17	29	27	28	18	14	13	5	2	180

These tables make clear, I am sure, that the first five tests may be used as an intelligence examination and that the results

will be nearly as dependable as they would be if the entire test were given.

Conclusion

1. That the probable success of first-year high-school pupils in the various courses can be predicted with a reasonably high degree of accuracy at the time they enter high school by the use of intelligence tests.

2. That the intelligence test clarifies the teachers' problems.

3. That the intelligence test offers the best criterion for the organization of class groups.

4. That the intelligence test affords guidance to teachers in the distribution of scholarship grades.

5. That the intelligence test, because of its stimulative effect on the individual, is a powerful agent of motivation.

6. That the results of intelligence tests must be used judiciously or otherwise great harm may be done.

7. That the Terman Group Test is well adapted to high-school freshmen.

8. That the last five tests of the Terman examination may be omitted without materially affecting the result.

9. In general, that the application of intelligence tests to first-year high-school classes is practicable and that unless we make use of them in improving the methods of teaching, in adapting the courses of study to the needs of the various groups, in guiding pupils in their program-making, in stimulating the individual to a greater realization of his possibilities, and in arousing in him a desire to measure up to those possibilities, we shall fail far short of our opportunities in these things.

Editorials

METRON

In July, 1920, a new international review called *Metron* was launched in Italy under the editorship of Doctor Corrado Gini, Professor of Statistics in the University of Padua. The second number appeared last December. This periodical is to appear as a quarterly and will amount to seven or eight hundred pages per year. It is published by the Industrie Grafiche Italiane at Rovigo (Veneto), Italy at fifty lire. At the rate of exchange at this writing this amounts to $2.04. This is indeed the time to buy foreign books.

As a literary venture *Metron* is unique. It accepts original articles in Italian, French, German, or English and prints them in the language in which they are written. Moreover, it is not exclusively or even particularly interested in a special field of knowledge. It is interested rather in the method by which authors investigate and report on problems, no matter what field they may represent. The one requirement is that papers shall represent a statistical treatment. Articles are also accepted on abstract methods in statistics. It is in this sense only that *Metron* may be said to have a specialty.

It is quite evident that the development of statistical inquiry has proceeded to the point where it is being erected into a field of knowledge if not into an independent science. This development has been going on for some time, and the publication of *Metron* is by no means the only evidence of it. For example, in 1913 Doctor Hugo Forscher wrote an important book on the theory of statistics which he entitled, "Statistical Method as an Independent Science."[1]

Originally, as the name implies, statistics had something to do with the state. It was not necessarily quantitative in character. As questions of state came to be more and more closely considered by an ever-increasing number of people, the necessity for a method of writing and reporting in a fixed language forced the adoption of numerical terms. Thus a rude procedure

[1] Forscher, Hugo. *Die statische Methode als selbständige Wissenschaft.* Leipzig: Veit and Company, 1913. 365 pp.

of a kind which we should now call statistical was built up. The application of this method in other fields naturally came about as the need for a rigorous treatment of their phenomena was felt. Thus we have social, economic, vital, and actuarial statistics, and later biological, psychological, educational, and financial statistics. There is now scarcely a field of knowledge in which statistics is not employed.

Of course, the field in which we are most interested is education. In it statistical methods have developed with a rapidity which has quite outstripped the ability of workers in education to comprehend it. Yet the time is coming when one can no more be a competent student of education without a working knowledge of statistics than one can be a competent actuary or economist without such knowledge. It is therefore altogether appropriate that a science of statistics should be recognized. It is not nearly so important that it be called a science as it is that it be accorded a definite position and appreciated as useful for its own sake.

The first number of *Metron* is a strong one. The first article (in French) is on "Statistical Method." It is a brief but well-organized general article on the subject. Important contributions are made by such distinguished statistical writers as Gini, Czuber, and Edgeworth. As indicating the wide range of topics, we may note the fact that one article (in English) belongs to the field of entomology and treats of the length of time which bees are away from the hive on each expedition, while another article (in French) deals with finance, and a third (in Italian) deals with horse-racing. This last article by Professor Gini is an interesting comparison between the success of book-makers at Rome and Milan as indicated by the relation between the actual order of finish of the horses and the order as expected according to the betting. The Roman members of the fraternity were more successful in picking the winners.

Our personal interest in a publication such as *Metron* has perhaps induced us to give it more space in these pages than our readers will consider appropriate. If we felt justified in devoting still more space to it, we might proceed to demolish our imaginary opponent; in other words, having set up our man of straw, we might proceed to knock him down. We feel quite sure that *Metron* is important as a statistical event. As such it should interest workers in all fields of knowledge to which statistics

apply. We are also sure that because of its polyglot character this publication is important for the graduate student in American universities because of the fine training it will enable him to obtain in the reading of foreign languages. We hope *Metron* will justify the fondest hopes of its founders. We admire their enterprise, and we wish them a full measure of success.

B. R. B.

THE LAW OF THE SINGLE VARIABLE

In the construction of educational tests there has been too little formulation of fundamental principles and too much mechanical application of statistical procedure. When, therefore, the author of a test accompanies the account of its derivation with a careful statement of the fundamental principles upon which she believes tests should be constructed, these principles deserve the thoughtful consideration of those interested in the construction and use of educational tests. Mrs. May Ayres Burgess in a monograph entitled *Measurement of Silent Reading* has formulated certain principles which she considers fundamental. These are epitomized in the title, "The Law of the Single Variable."

The three variables of a pupil's ability, or more strictly speaking, of his performance, are: (1) quality of performance; (2) level of difficulty on which it is given; (3) amount of work done within a fixed time, or rate of work. These variables are recognized as being complex but they are considered to include all factors which affect a pupil' performance.

Performances which depend upon a single variable are illustrated by the author in the field of athletics. In the high jump the variable is the heighth of the bar cleared. It represents the difficulty of the performance. The quality and the rate are either negligible or constant. In the races the rate is the variable, both difficulty and quality are either constant or eliminated. In shooting at a mark, quality is the only variable considered. In all of these cases the ability of a contestant is measured in terms of a single variable. The variable which is used is determined by the nature of the ability.

In applying this law of the single variable in the field of educational measurements the author states that it "consists of distinguishing the possible controlling, varying factors; devising means of holding them all constant save one; and measuring that

one." Correspondence with the author reveals that this statement was not intended to mean what it appears to mean. It appears that the law of the single variable was not intended to be interpreted as meaning that two of the variables must always be kept constant for all pupils while the third was measured. It should be interpreted to mean that pupil performances depend upon three variables, or have three dimensions. In describing a performance these three variables must be recognized. If a variable is not constant for a group of pupils its variation must be recorded for each child in the process of testing and when scores are interpreted they must be on the basis of a single variable. For example, if children are to be compared with reference to their rate of work it must be shown that both the quality and the difficulty of the work done were the same for all pupils. "What the law of the single variable does not permit is the attempt to compare combinations of the three variables in unknown and varying amounts."

It is obvious that the law of the single variable should not be interpreted to mean that when a test is given all pupils should be forced to give a performance which is constant with respect to two of the three variables. The characteristics of the abilities which children acquire are not restricted to a single variable. When a group of pupils is performing in a given subject-matter field in the way which is most natural for each, large individual differences are exhibited with respect to these three variables, particularly with respect to rate and quality of work. This is true even in the case of groups of pupils which have received the same instruction. It is reasonably easy to control the difficulty through the selection of the content of the test. The control of the rate of work, or the quality of work is not so simple. In certain fields it may be shown that one of these factors has little effect upon the performance. For example, in spelling the rate of work has little effect upon a pupil's performance, provided the time allowed is sufficient for all pupils to write the words. In some fields the rate is an unimportant variable and can be neglected. This would be true in the case of painting or drawing when the products are real works of art. In these cases the ability of the performer depends upon the single variable of quality.

If a variable is controlled by an arbitrary procedure which produces unnatural conditions it appears likely that the ability

which functions is modified, or factors are introduced into the testing procedure that produce the same result upon the performance. For example, if in the case of handwriting all pupils of a group were forced to write at a fixed rate, those who were accustomed to write more rapidly or more slowly than this rate would exhibit abilities different from those which normally function in their handwriting. To the extent that this is done the performance secured would not be a true index of the ability of the pupil

Although the author fails to make it clear in her monograph the "moral" of the law of the single variable appears to be that measuring instruments must explicitly recognize the possible existence of the three variables, or dimensions, of a pupil's performance. These must be described separately if accurate interpretation of the scores is to be possible. If a variable is omitted in a pupil's score it must be shown to have been constant in its effect upon the performances of different pupils or to be socially unimportant. This is a principle which has been sadly neglected by many of the makers of educational tests. In addition to yielding measurements which are likely to be erroneously interpreted, instruments in whose construction this principle is violated also lead to implications that are not in agreement with our educational objectives. Grade norms for a test are generally interpreted as standards or goals to be obtained in the respective grades. An instrument which measures the pupil's ability in terms of the most difficult step reached therefore implies that our objectives in that field are to prepare the pupil to do more and more difficult things. We should teach pupils to do not merely difficult things but things that are socially useful. Instruments which fail to measure the rate of work in fields where it is an important variable are also open to criticism. They do not tell the whole truth about a pupil's ability.

Dr. Burgess has rendered a real service in formulating the principles embodied in the law of the single variable. It is unfortunate that they are not stated in such a way that misinterpretation would not be likely. Whether or not future makers of tests are prepared to accept the law of the single variable it is hoped that they will feel its incumbent upon them to formulate carefully the principles of measurement upon which the construction of their instrument is based.

W. S. M.

Reviews and Abstracts

E. H. CAMERON, *Editor*

BURGESS, MAY AYRES. *Measurement of silent reading.* New York: Department of Education, Russell Sage Foundation, 1920. 163 pp.

In this monograph the author describes the derivation of a new type of scale for measuring ability in silent reading, which is called "Picture Supplement Scale (P. S.—1)." Supplementary to this account the author summarizes certain general considerations relative to the theory of educational measurements. The Picture Supplement Scale is the product of a series of carefully planned experiments. The chapters devoted to the theory of educational measurement are also evidence that much thought was given to the fundamental principles on which the construction of educational measuring instruments should be based. In both cases the author has made a notable contribution to our literature in this field.

The wide-spread interest in educational measurements has resulted in the derivation of an exceedingly large number of measuring instruments within the past few years. For the most part the authors of tests have accepted without critical consideration, methods of test construction which were originally developed in a single field and have applied them without any critical analysis to the field in which they happen to be interested. In the midst of such work it is refreshing to read the account of the derivation of the Picture Supplement Scale. Statistical tables and the description of an elaborate statistical procedure are conspicuous by absence. The major portion of the monograph is devoted to a critical consideration of the principles which should govern the construction of tests with particular reference in the field of silent reading. The central theme of the part devoted to the theory of educational measurements is summarized in the caption "The Law of the Single Variable." The writer comments upon this law editorially elsewhere in this issue of the JOURNAL OF EDUCATIONAL RESEARCH and for that reason only brief reference will be made to it here.

The function of the "Picture Supplement Scale" is to measure the amount of printed material which the pupil can read within a given time. In respect to the type of reading which is measured the author claims that it is "a test of careful reading." The scale consists of twenty exercises equal in difficulty and printed on one side of a 12 x 19 sheet. Each exercise consists of a short paragraph and a picture. The paragraph is partly descriptive of the picture and in part consists of directions for drawing a supplementary picture. The test of the pupil's reading of the paragraph is the drawing of the supplementary picture. The pupils are allowed five minutes to do as many as they can. Directions for scoring are liberal. Any sort of a drawing which shows that the pupil followed instruction is to be counted as correct. The scale is recommended for use in grades III to VIII.

In the derivation of this scale the author identified twenty-five factors which influence a pupil's performance in silent reading. An attempt was made either to eliminate or to control twenty-four of these factors, leaving only the amount read as the variable factor. The ability to draw the pictures required (difficulty of action

demanded) is announced as being held constant but one naturally questions its constancy. Some of the drawings to be made are very simple. For example, in Exercise 6 a pupil is asked to cross out a portion of a picture. In other exercises the drawing required is more complex. In Exercise 4 the pupil is to draw a picture of three feathers. No instructions are given to the pupil concerning the quality of the drawings required. One would, therefore, expect to find that some pupils would attempt careful drawing while others would make very hasty sketches.

The instructions for administering the tests are meager and do not specify the exact explanations to be given to the pupils. The pupils are given no preliminary exercises to acquaint them with the nature of the tests. One would, therefore, expect the administration of the tests to lack objectivity. One is also inclined to question the objectivity of the scoring, even though the directions state that any sort of a drawing which shows that a pupil has followed instructions is to be accepted as correct. In fact one state bureau which is distributing the Picture Supplement Scale has found it necessary to issue detailed directions for scoring the exercises.

The pupil's score is the number of exercises done correctly in the five minutes allowed for the test. The author considers that both the difficulty of the exercises and the quality of reading required are kept constant. The uniformity of difficulty is secured by use of exercises for which the same percent of correct responses is in general obtained. The quality of the reading is considered to be kept constant because the pupil is given credit for only those exercises which he answers correctly. Hence, according to the author the score is a measure of the amount of reading which the pupil has done in five minutes or, in other words, his rate of reading.

It is likely that the number of exercises done correctly furnishes the best single numerical description of a pupil's performance on this test. However, this score is not a measure of his rate of work. It is a combination of rate of reading and quality of reading. This combination is not the same for all pupils. For example, a pupil may make a score of 10 doing ten exercises with 100 percent accuracy. Another pupil may make a score of 10 by doing twenty exercises with 50 percent accuracy. The rate of reading is not the same for these two pupils. Neither have they read with the same quality.

The pupil's score in terms of the number of exercises done correctly is to be translated into a score on a percent scale. This scale is different for the different grades. For example, if a pupil is in the fourth grade ten paragraphs done correctly entitle him to a score of 68. If he is in the seventh grade his score is 50. No real advantage is gained by this translation and it will doubtless be confusing to many.

The reliability of the scale was studied by having it given twice to a few small groups of pupils. The reliability of coefficients calculated for these groups closely approximate those for a number of other reading tests.

The Picture Supplement Scale represents an ingenious and serious attempt to construct an instrument for the measurement of silent reading ability. It deserves a place among our best reading tests. If the present form were supplemented by more detailed directions for administering it and by the addition of a few preliminary exercises, it would be well adapted for general use, particularly in the lower grades.

W. S. M.

CUBBERLEY, ELLWOOD P. *The history of education.* New York: Houghton Mifflin Company, 1920. 849 pp.

CUBBERLEY, ELLWOOD P. *Readings in the history of education.* New York: Houghton Mifflin Company, 1920. 684 pp.

Professor Cubberley has, through the exceedingly useful character of his educational work, secured a very large following who will welcome the publication of these two important contributions to the history of education. Professor Cubberley is fortunate in having attained to an almost unique position in the educational world. Although for many years he has been directly connected with university work he has kept in the closest touch with the developments of educational administration and, both through his writings and his work in various cities as an educational surveyor, he has personally contributed powerfully to the general acceptance by administrators of standards in educational administration. Having retained this intimate and constructive attitude toward education over so long a period of years one expects to find his treatment of the history of education modified by this point of view.

This expectation is realized, as he himself points out in his preface. It is difficult for one who is continually surveying the work of public school systems not to see the very close connection which always exists between a civilization and its type of schools. The development of educational theory, as such, is forced to the background while the consideration of the actual efforts in education is correspondingly emphasized. It is probably due to the rather general realization of the indefiniteness in the relation between educational theory and the solution of present educational problems which has caused during the last few years the sharp decline in popularity of the traditional courses in the history of education. At any rate, we have in these two books the concrete efforts of a man who is primarily interested in the effort of modern society toward realizing itself effectively through improvement in school organization, to prepare a history which would treat education as a phase of the general development of civilization. That he has accomplished this in a way which will win the hearty approval of the school men of the United States almost goes without saying.

Were one inclined to be over-critical, he might question the wisdom of including within the pages of the text so much that is found in any well-written survey of history. He might ask himself whether the book could not easily have been made much smaller and yet not be reduced to any degree as a presentation of the development of education. I found myself continually asking this question; and yet, when I specifically attempted to cut out topics that primarily belonged to general history I hesitated, realizing that in spite of the greatest efforts of teachers of history, much of this information might be so poorly mastered by the student that the real significance of the educational topics would be lost without this other presentation. So I have finally withdrawn my criticism believing that from the teaching standpoint Mr. Cubberley justifies the inclusion of this material.

All of Mr. Cubberley's texts are unusually readable. These volumes are no exception to the rule. I predict that many schoolmen will, from the motive of pleasure alone, read the text from cover to cover.

The *Readings in the History of Education* have been carefully selected and are so arranged as to follow the chapter arrangement of the text. In this way much that might otherwise have had to be included in the presentation is immediately available in the second volume. So well has this collection of readings been compiled that none

but a specialist in the history of education would feel much need for access to source not found within the covers of this book.

Professor Cubberley follows the customary organization of material as found in texts now generally used but never loses sight of the point of view stated in the subtitle of the volume "Educational practise and progress considered as a phase in the development and spread of western civilization." He divides his study of western civilization into four parts: "The Ancient World," "The Mediaeval World," "The Transition from Mediaeval to Modern Attitudes," and "Modern Times." He, however, appreciates fully the vastly greater significance to the student of education of the modern period and devotes approximately half the book to part 4. Students of history will doubtless question the propriety of dismissing the educational contribution of Greece, Rome and early Christianity with a paltry hundred pages. Possibly one might query whether complete omission might not be justified with better grace than the cursory mention which, frequently, is all that is included. I, personally, however believe that the distribution of emphasis is well justified. Unless prerequisites are insisted upon, complete omission produces with many readers distorted perspective, while cursory mention suggests the influence upon modern conditions which is all for which the author is striving.

Certainly he is justified in assuming that the student or reader is not completely ignorant of ancient and mediaeval history. Therefore he is equally justified in assuming that in this volume the chief purpose is to show the vital connections between the significant development of the civilization of Greece and Rome and the educational conditions of those ages, and that in turn Christianity not only furnished the only institution which could serve as a foundation for modern civilization but conserved certain fundamentals both of civilization and education without which the task of reconstruction would have been far more complex and difficult. Cubberley's treatment of the significance to modern education of the revival of learning, of the Protestant revolts, of the development of scientific inquiry and of the scientific method, while conforming to that usually found is clear and convincing and dramatically real to the average reader.

School superintendents and principals will be especially impressed with the discussions centering around the abolition of privilege, the rise of democracy, the new theory of education, and the struggle as a result of which the state takes over the school.

Those who have read his *Public Education in the United States* will find much that is duplication of the earlier work. This is probably inevitable. Professor Cubberley could not afford to omit important and pertinent topics merely because he had already discussed them in another volume. At the same time, those who read all of his writings may well be pardoned if they regret that he did not see his way clear to vary more his presentation of these topics. The same feeling has occurred to me more than once in reading *Public Education in the United States* and comparing it with his *Public School Administration*. Topics in school administration although legitimately and necessarily found both in a history of education that professes to include the educational movements that have had their best exemplification in the United States and in a special treatment of education in the United States surely can be developed from somewhat different angles if they frankly make their appeal to the same group of readers. However, this is really a very minor criticism. Most of us will profit by rereading these topics and those who use only one of the volumes receive the point of view in each case that the author is earnestly desiring to impart.

Again the scientific historian might question the propriety of including so much that is so recent that it is still in the process of making. I, personally, however, find in this one of the chief charms of the book. Cubberley makes our history of education a real pulsating thing. He shows how the older movements affect present conditions and how just as in the past our educational organization is a method through which our people become better adjusted to the conditions under which they live and which they must in turn affect.

I commend these two volumes to teachers of the history of education in our departments of education in universities and to teachers of the history of education in our normal schools as handbooks admirably designed to give a student of this subject a point of view towards education which is eminently sane and constructive. Too many teachers and even administrators fail to realize the cause of many of our present educational problems. Studies of this character develop the power to see the relationship between a given problem and the given social, political, and economic conditions. In my judgment, therefore, a course in the history of education from this point of view is worth while for all students of education to take and I would not be at all surprised to see as a result of the publication of these volumes a very real increase in the influence of the history of education as a vital division in our departments of education.

C. E. CHADSEY

University of Illinois

CUBBERLEY, E. P. *The history of education.* Boston: Houghton Mifflin Company, 1920. 849 pp.

In his preface, Professor Cubberley tells us that he has "not tried to prepare another history of educational theories," and that he has "omitted reference to many theorists and reformers." His dismissal of Plato and Aristotle in a few sentences, and without giving even the briefest sketch of their schemes of education, is equivalent to omission of reference. The reader is told that Quintilian was "the foremost Roman writer on educational practice," and that his *Institutes of Oratory* gives "a detailed explanation of the old Roman theory of education at its best," but is left to wonder what "the old Roman theory of education at its best" was like. "For such omission I have no apology to make." *Ipse dixit.* Very well! This review, then, to borrow the author's phraseology, will be "less concerned . . . . with the educational and philosophical theories advanced by thinkers . . . . than with what was actually done." However, it should be observed that a well-balanced History of Education, or History of Civilization, should contain liberal reference to "theories."

As one reads the text he is conscious of passing through many areas of condensation and rarefaction. The reviewer questions the advisability of subjecting the casual reader, or immature student, for whom the work is designed, to so many changes of pressure.

The preface, and the book itself, suggest that it may be more profitable to search for errors in the statement of "facts," than in the presentation and evaluation of "theories." And their "name is legion."

On page 7, in the context having reference to the opening of the "Christian era," appears the statement that "Greek was forgotten." The fact is, Greek was not forgotten at that time. The following page contains the information that "the study of Greek and Hebrew" was "revived" during the "Italian Revival," and on page 95, in a

section devoted to the "Rejection of pagan learning in the West," (fourth-sixth centuries) we read that "the Greek language was forgotten, and was not known again in the West for nearly a thousand years." This incorrect impression is enhanced by the statement (page 424) that "Greek . . . . was restored to the western world" during the period embraced by the years 1333–1433. The study of Greek was not unknown in Sicily and Italy in the twelfth century, and the "study of Greek and Hebrew" occupied a fairly prominent place in the thirteenth century. John of Capua, Raymond Lull, Siger de Brabant, Guillaume Bernardi di Gaillac, and Arnauld de Villaneuve knew Greek and Hebrew. The Dominicans made provision for instruction in these languages in the thirteenth century, and the Council of Vienne, 1312, decreed that the Eastern tongues be taught at Rome, Bologna, Salamanca, Paris, and Oxford. It is incorrect to say that "during the Middle Ages . . . . none could read it" (Greek). The "study of Greek" continued, without interruption, throughout the Middle Ages.

A legend that no longer obtains among those who have done intensive research in the field of mediaeval literature is continued in such expressions as the following: "the long intellectual night of the Middle Ages," "the long intellectual night of the mediaeval period," "the long period of barbarism and general ignorance," and "the long period of intellectual stagnation."

True enough, during the seventh and eighth centuries, "Many of the priests were woefully ignorant," but many have been "woefully ignorant" ever since that day. An influential few, however, were not so ignorant. And one must take *cum magno grano* the statement that "the Latin writings of the time contain many inaccuracies and corruptions which reveal the low standard of learning even among the better educated of the clerical class." The standard of the educated clergy was high; it was "low" only among the uneducated clergy. Would it be proper to conclude that a "low standard of learning" obtains among present-day writers, of college texts even, because their "writings . . . . contain many inaccuracies and corruptions"? Certain canons of historical method should be observed in our evaluations and interpretations.

The conclusion (p. 263) that "Mediaeval education . . . . prepared for the world to come; not for the world men live in here" is too sweeping. Only the narrowly religious education of the period "prepared for the world to come." Secular education, whether for the nobles and the well-to-do, or for those of "low degree," did prepare "for the world men live in here." "The world of the mediaeval monk and the Scholastic" (p. 279) was not the whole world of the Middle Ages, by any means.

At one point (p. 155) we read that theology was "the one professional study of the whole middle-age period," and at others (pp. 199, 211) that medicine and law were "professional subjects" also.

For the convenience of the reader of this review five consecutive sentences, that occur under the title "Results of their work" (the work of the Schoolmen, p. 192), are here reproduced. "The work of the Schoolmen was to organize and present in systematic and dogmatic form the teachings of the Church. This they did exceedingly well, and the result was a thoroughgoing organization of Theology as a teaching subject. They did little to extend knowledge, and nothing at all to apply it to the problems of nature and man. Their work was abstract and philosophical instead, dealing wholly with theological questions. The purpose was to lay down principles, and to offer a training in analysis, comparison, classification, and deduction which would prepare learned and subtle defenders of the faith of the Church." Evidently the author wishes to give the impression that the Schoolmen dealt "wholly with theological questions."

What could be further from the truth! Scholasticism was interested in every phase of intellectual activity. Only part of its work was that of "reorganizing and systematizing theology."

A rather striking area of condensation contains the following, in two sentences: "Luther's . . . . translation (New Testament, 1522) . . . . virtually fixed the character of the German language. Calvin's *Institutes of Christianity* (French edition, 1541) . . . . fixed the character of the French language, and Tyndale's translation of the New Testament (1526) . . . . fixed the character of the English tongue." Sweeping, to say the least! And these statements occur without qualification.

On one page (258) we find that Magellan "circumnavigated the globe" in 1519–21, and on another (424) that his "ships rounded the world" in 1515–18. A similar discrepancy occurs in connection with the dates of publication of Rousseau's *Social Contract*, and *Émile*. "In 1752 appeared both his *Social Contract*, and *Émile*" (483). Page 508 indicates that they were published in 1762. In another instance (p. 525) the author says that "In 1799 . . . . Jefferson tried unsuccessfully to secure the passage of a comprehensive bill . . . . for the organization of a complete system of public education for Virginia." If this is a typographical error for 1779, it should have been corrected in a footnote to page 526, and in the *Readings*, page 427. Again, an incorrect impression might be gained from the statement (p. 524) that "New York in 1787 created an administrative organization known as the University of the State of New York." But, with the author's *Public Education in the United States* at hand, the reader is fairly safe; for here (p. 258) he learns that the University of the State of New York was "established in 1784, and organized in its permanent form in 1787."

The author quotes, without comment, on page 369, the following, from Draper: "All the English schools in the province (of New York) from 1700 down to the time of the Declaration of Independence were maintained by a great religious society . . . . called the Society for the Propagation of the Gospel in Foreign Parts." As a matter of fact, but very few of the schools in New York during this period were maintained by the S. P. G. And it is equally incorrect to say (p. 521) that "In the Middle Colonies, where the parochial-school conception of education was the prevailing type, the school remained under church control until after the foundation of our national government." Most of the schools of the Middle Colonies, in the eighteenth century, were private-venture schools, and not "under church control." Again, a faulty implication is contained in the assertion (p. 520) that "By the close of the colonial period the new American Academy (p. 463), with its more practical studies, had begun to supersede the old Latin grammar school." Schools offering these "more practical studies" far outnumbered those of the Latin Grammar School type long before "the close of the colonial period."

The average instructor, as well as the average student, accepts the printed word. If care is observed, to point out certain errors, the text is usable. Its scheme of organization is good. It supplements admirably the work that has been done by Graves and Monroe.

ROBERT F. SEYBOLT

University of Illinois

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Walter S. Monroe, University of Illinois, Urbana, Illinois.

Educational Research in Oklahoma

A letter from Henry D. Rinsland, Director of Department of Research and Guidance, Ardmore (Oklahoma) City Schools, gives the following information concerning educational research activities in his state.

At the invitation of County Superintendent Mrs. M. O'Daniel Rinsland, a survey is being made of the public schools of Johnston County. This is the first county survey in Oklahoma. The members of the survey staff are from the University of Oklahoma and two of the state normal schools. The survey has been organized in seven divisions as follows:

Division 1. Organization and Administration
Division 2. Buildings and Equipment
Division 3. Attendance and Enrollment.
Division 4. Instruction, Course of Study
Division 5. Teacher Status
Division 6. Tests and Measurements
Division 7. Finance.

Mr. Rinsland also reports the extensive use of intelligence tests in the public schools of Ardmore under his supervision. He has also spent some time in Henrietta supervising the giving of educational and intelligence tests in that city.

Bureau of School Service University of Kansas

At the University of Kansas, Dean F. J. Kelly of the School of Education conceived the idea of a bureau whose function would be primarily service rather than research. At the present this Bureau of School Service is under the direction of F. P. O'Brien. A recent letter from Professor O'Brien gives the following information concerning the work of this bureau.

"The Bureau has directed a survey of the curriculum and building problems of the Lawrence school system. The report was completed early in January and will soon appear in printed form. The Bureau is at the present time cooperating with the schools in cities of the first, second, and third classes, in making a study of the teacher-salary situation in the schools of the State. The report is nearing completion and will soon be distributed to the school systems in this State so that it will be in their hands before the time for the re-employment of teachers for the coming year. The Bureau is also conducting a survey of the school situation as it is found in several of the school districts in an adjoining county of Kansas. Some of these districts are in a condition of distress due partly to the fact that several small high schools are maintained where one should be adequate. The financial burden of the present arrangement is crushing for certain districts. They made an appeal thru the County Superintendent to this

Bureau of the University for some professional assistance in advising them what to do. That whole section of the county is now anxiously awaiting the report of our investigation.

"It is evident that such undertakings are chiefly of the service type but the Bureau has also been promoting Research Work in a smaller way by assisting and encouraging students interested in experimental or research study."

Latin Tests by a Latin Specialist

Dr. Evan T. Sage, Professor of Latin of the University of Pittsburgh, is developing some Latin tests. In a recent letter he says, "We are going on this assumption for the time being that in addition to certain qualities which we may not be able to measure a student needs some knowledge of forms, vocabulary and syntax. There are of course tests in existence to measure all these things separately. But as the concrete aim of Latin teaching is ability to translate it should be possible to make a translation scale which will measure achievement in all these things at once and also serve as a basis for investigation of the other elements involved in successful translation. The tests of this sort which are now in existence seem to be open to the objection that they are not based on the standard vocabulary and standard syntax. Unfortunately inflections have not yet been standardized though we are working on that problem."

It is quite within the bounds of reason that a Latin test by a professor of Latin may take precedence over every such test now available. We have long thought in the testing of abilities in subjects taught in secondary and higher educational institutions the experts in these subjects should be heard from. It takes something more than a mere scale maker to devise acceptable instruments of this sort. We hope Professor Sage will carry his investigations to a successful conclusion.

A Bulletin on Silent Reading

We are in receipt of a "Bulletin on Silent Reading" from Superintendent F. S. Camp of Stamford, Connecticut. It is addressed to his principals and teachers. It is one of the best we have seen. Both silent and oral reading are divided into intensive and extensive activities, each of which is defined and illustrated.

"Silent intensive reading" is largely study. "Silent extensive reading" is reading for pleasure and individual interests. "Oral intensive reading" includes in the early grades the conventional mechanics of reading and in the upper grades the study of masterpieces and "group study reading," i. e., reading in which all children have the same texts before them and are actuated by a common purpose. "Oral extensive reading" occurs when one person reads to a group of listeners.

The bulletin gives for each of these four types of reading activity a suggested proportion of time to be utilized in each of the eight elementary grades. The shift of emphasis from grade to grade, so far as the program may show it, is thus clearly set forth in the table on page 70. The entries indicate the number of tenths of the entire reading time in each grade.

Superintendent Camp discusses the question of reading from these four points of view and provides definite means by which the principals and teachers may check up their success in meeting the objectives he sets up. Several blanks for recording the amount and character of the reading done by children are provided. This sort of a bulletin cannot fail, especially when followed up by the supervisory staff, to produce *measurable* improvement.

SUGGESTED TIME VALUES, TEN REPRESENTING THE BASE FOR READING

Type of Reading Activity	Grades							
	I	II	III	IV	V	VI	VII	VIII
Silent intensive reading	2	1	a	a	a	0	0	0
Silent extensive reading	0	1	2	3	5	6	7	7
Oral intensive reading	7	6	5	4	3	2	2	2
Oral extensive reading	1	2	3	3	2	2	1	1

[a] Large time values here in grades where the child must *learn* from the printed text; but as it is not considered as falling under "reading exercises" in the ordinary sense, no ratio numbers are assigned.

Schoolmen's Week University of Pennsylvania

Schoolmen's Week at the University of Pennsylvania was held Thursday, Friday and Saturday, April 7–9. There were 1,885 persons registered while 8,565 attendances were noted in 31 programs. The number of registrations and attendances were both larger than in any preceding year, the previous highest records, 1,392 and 3,802 respectively, having been made in 1920.

Among the speakers from outside the state were, J. H. Kirkland, Chancellor, Vanderbilt University, Nashville, Tennessee; Frank Aydelotte, President-elect of Swarthmore College, now Professor of English, Massachusetts Institute of Technology; John W. Withers, former Superintendent of Schools of St. Louis, now Dean of School of Education, New York University; Walter S. Dearborn and John M. Brewer, Harvard University; William S. Gray, School of Education, University of Chicago; Professor N. L. Engelhardt, Otis W. Caldwell, and Miss Fannie Dunn, Columbia University; and C. J. Galpin, Specialist in Farm Life Studies, Department of Agriculture, Washington, D. C.

Social and recreational features took the form of visitation of the campus and buildings, including the museum and library. The guests of the university were given luncheons on Thursday, Friday, and Saturday at noon, and on the evenings of Thursday and Friday. On Saturday over 650 luncheons were served, a reception was tendered by the Acting Provost, Doctor Josiah H. Penniman after the Friday evening program. Several hundred accepted the invitation of the Athletic Association to witness the baseball game between Pennsylvania and Swarthmore, Saturday afternoon. On Thursday night two hundred superintendents and supervising principals were given stage seats at the Academy of Music and the Metropolitan Opera House where the American Legion gave a program of unusual interest for the purpose of stimulating the spirit of Americanism.

It was generally considered the most successful meeting yet held. This result was due in large part to the faithful and cordial cooperation of the members of the General Committee, the Reception and Registration Committees, and the Advisory Committeemen, and to the support given by State Superintendent Thomas E. Finegan and City Superintendent Edwin E. Broome and their assistants. The executive officers

were the same as in previous years—Harlan Updegraff, Chairman, General Committee; Arthur J. Jones in charge of Secondary School Conferences and LeRoy A. King, Secretary, General Committee.

Equivalence of Forms I and II of the Burgess Picture Supplement Scale for Measuring Silent Reading Ability

Soon after the beginning of the present semester, the Burgess Picture Supplement Scale for Measuring Silent Reading Ability, Form I, was applied to about 1,000 seventh- and eighth-grade pupils with the following results:

Class	No. Tested	Paragraphs (Median)	Credit (Median)
VII–B	238	7	35
VII–A	263	8	38
VIII–B	242	9	41
VIII–A	219	9	38

It was expected that later in the semester Form II of the same scale would be applied for the purpose of securing a measure of the progress that is being made under present methods of instruction.

In order to get an idea of the practice effect of taking the test once, a group of high sixth-grade pupils was given Form I on April seventh. Under exactly the same conditions so far as it is possible to secure such a situation the same group was given Form II on the following day. The records of all pupils not present at both tests were discarded.

As in the case of the seventh and eighth grades, this group fell below standard, securing a median score of only 43. But to our surprise the median on the following day with Form II jumped to 64. Forty individuals made gains totaling 198 paragraphs, four neither lost nor gained and two lost a total of 2 paragraphs.

These gains were so notable that it seemed as if they could not be attributed entirely to practice effect. Therefore another group of 45 high sixth-grade pupils was selected and given Form II, followed on the second day thereafter by Form I. Instead of an improvement appearing as the result of experience there was manifested a decided loss. The median score dropped from 55 to 49. Twelve pupils gained twenty paragraphs. Nine neither gained nor lost and twenty lost a total of 55 paragraphs. This helped to conform our suspicions that the gains of group one were partly due to a difference in difficulty of the two forms and not altogether to practice effect.

To further satisfy ourselves as to whether or not Form II is as difficult as Form I, a third group of high sixth-grade pupils was given Form II and after a brief rest, Form I. The general results agreed with the facts already noted. This class scored 58 on Form II and after a few minutes rest fell to a score of 49 on Form I. Seven pupils made a total gain of 10 paragraphs, over the record made on Form II, seven pupils made the same score and eighteen pupils lost a total of 43 paragraphs.

Our brief experience with this test may be summarized thus: (1) Form I appears to be considerably more difficult than Form II; (2) inasmuch as all of our Form I

median scores are below standard and all of our Form II median scores are abov standard we may be justified in a suspicion that Form I standards may be a tride hig and that Form II standards may be too low. Of course, we claim for our results values beyond that of the proverbial straw that indicates which way the wind probably blowing.

Information concerning the experiences of others bearing on the main question here involved will be eagerly awaited. Meanwhile we shall be in doubt as to how t complete our undertaking to secure a correct measure of the semester gains made by our seventh- and eighth-grade classes.

H. C. Daley

Survey Department,
Highland Park Public Schools,
Highland Park, Michigan

The Illinois Examination

While there is much discussion both pro and con in regard to tests students of education, supervisors, and teachers may be interested in reading of the use of the Illinois Examination in Bureau County and the Princeton city schools in October, 1920. The writer has been actively engaged in working with this battery of educational tests in Bureau County since the sixteenth of last August. The examination in this county involved three hundred eighty-five teachers and about five thousand five hundred boys and girls. It is impossible to give any account of the Princeton city schools without discussing the work of the entire county.

County Superintendent George O. Smith of Bureau County is not only a firm believer that educational tests will be a big factor in the solution of future school problems but he is also anxious that his teachers be prepared to participate actively in the use of them. As a result the writer offered work on educational tests two periods a day in the August institute. The Illinois Examination and Teachers' Handbook were made the basis of instruction. Two sections of about one hundred teachers each were formed. The work of each section was similar. Consultation periods were also provided. The work was given under the following divisions:

1. Each teacher was given the test in exactly the same manner in which she was expected to give it.
2. Each teacher was taught how to score the papers using her own as a sample.
3. The interpretation of mental age and intelligence quotient in terms of ability to do work of the grade in which the child was located was explained.
4. The interpretation of achievement ages and achievement quotients was considered with reference to planning suitable remedies for unsatisfactory instruction.

After giving tests, teachers were asked to make out record sheets like the one below. These are on file for reference at any time pupil promotions are being discussed. The column marked "Health" is to be marked for each pupil with 1, 2, or 3. One stands for excellent health, two good health, and three poor health. The health statement is filled out by the school nurses. The remainder of the sheet is filled in for each child from his examination.

Many criticize tests because we do not follow up. In the case of Bureau County and the Princeton city schools the follow-up work was carefully planned and carried out. The entire county was divided into districts for the convenience of the teachers.

BUREAU COUNTY PUBLIC SCHOOLS

ILLINOIS EXAMINATIONS GIVEN October 7, 1920

ool Lincoln (Princeton) *Teacher* T.V.Fisher *District* 115

Names of Pupils (alphabetically)	Health	Grade	Chronological Age yrs. mos.	Mental or Achievement Ages					Intelligence or Achievement Quotient				
				Intelligence	Arithmetic	Reading			Intelligence	Arithmetic	Reading		
						Average	Comprehension	Rate			Average	Comprehension	Rate
tbur Rhoek	3	5	14	10	7	10	8	12	75	71	100	80	120

A called teachers' meeting was held in each district. At the meeting the test was carefully reviewed and the typical type errors called to the attention of the teachers. First of all, teachers were urged to secure a textbook[1] that would deal with the subject of remedies in detail. The arithmetical errors we arranged as to type. It was found that in arithmetic pupils were generally weak in the fundamental operations. To improve this condition drill cards[2] in arithmetic were secured. The teacher was then instructed to go over the test papers carefully and let each child work on those drills in which he did not make satisfactory scores. Each teacher is also provided with a sheet of problems showing the different abilities required in the fundamental operations. This same idea was carried out in all phases of the work in arithmetic. It will be readily seen that in this way the pupil received individual attention and was not required to drill or work on material that he could handle successfully.

The grade medians for the Princeton city schools are approximately equivalent to the grade standards in the case of general intelligence. This means that the boys and girls in the different grades of the Princeton city schools have, on the average, about the same capacity to learn as that exhibited by the average corresponding grade in other places. This does not mean that every boy or girl is up to standard with respect to general intelligence. We have some pupils who differ from it by as much as three years. In grade VII we have mental ages ranging from nine to eighteen years. The least range is found in the third grade.

The median scores for silent reading were up to the standards. This verifies several other tests that have been given in the last two years to the pupils of the Princeton city schools and is probably due in a measure to the methods used in the teaching of the subject.

[1] Monroe, W. S. *Measuring the results of teaching*. New York: Houghton Mifflin Co., 1918.
[2] Clark, M. G. Arithmetic Habituated, Series A.

The median scores in arithmetic are not satisfactory to the teachers nor are they up to the standards of the examination. There are several known causes back of this.

1. The test involved, in certain grades, operations that are not introduced in the Princeton city schools until later in the life of the child.

2. Previous to the present school year the arithmetic text was not satisfactory nor practical.

3. The instruction was not suitable as it was based entirely on the text in use.

The first of these causes cannot be overcome in our course of study. The second has been remedied by a change in arithmetic textbooks. The third is being worked out as rapidly as possible. At present some improvement has been shown. The superintendent has gone over the pupils' papers with the classroom teacher and the errors have been checked. Drill is being given each day on the basis of the individual needs.

The use of the Illinois Examination in Bureau County and the Princeton city schools has suggested the following observations with reference to its value.

The Illinois Examination is simple enough in its operation to make it a valuable aid to both rural and city schools. It is very easy to administer.

Teachers who have used it are inclined to diagnose more carefully all cases of failure to do good school work.

Teachers, upon discovering that this work is constructive and not for the purpose of finding fault with teachers, cooperate very willingly.

While the examination is standardized for the group and was not used for promotional purposes it has designated certain children as worthy of larger opportunities to make progress.

A better understanding of school situations is developed between teachers and supervisors.

The examination forms a preliminary basis for the study of gradation or grouping by mental rather than chronological ages.

It will be no small factor in making teachers realize that there is a scientific side to teaching that cannot be neglected.

CARLETON BLOSE SMITH

Superintendent, Princeton, Illinois

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

DOES EDUCATION PAY FOR ITSELF?

Recently at an educational meeting, a university professor spent forty-five minutes in elaborating the proposition that the public should be taxed to support only those forms of education which clearly return full economic value to society and that the individual should pay in full for those other forms of education which do not make such returns. He upheld the public elementary schools in general though insisting that their course should be changed in such a manner that they might well be called schools of citizenship. He suggested that perhaps these schools should carry the youth to about the age of sixteen by which time the principles of government and economic law would be so well grounded that the children so taught would hence forth be good citizens.

He indorsed public support for the graduate schools with emphasis upon the research laboratories, holding that they had already demonstrated that they return full value for the investment made by public taxation. Concerning the two remaining phases of education which are now largely supported by public taxation, secondary and undergraduate higher, he had serious doubts of the propriety of continued support. He expressed these doubts in such a manner that his hearers recognized that he had no doubts in his own mind. He was quite sure that the public could not continue the present trend of publicly supported educational opportunity for ever-increasing numbers, and that high-school and college education does not pay for itself in returns to society at large; perhaps not even in returns to the individual receiving it.

I am not concerned in this column with a debate of the subject. I do not conceive that this is the place or the time for either a rebuttal of his argument or a further elaboration of the correctness of his position. But I do believe that the thought expressed may legitimately be used as a basis of serious thought and comment apropos of the work of members of bureaus of educational research.

Let us face the question: not does education pay for itself but does my work pay for itself? Is the city or state for which I am working getting an adequate economic return for the expenditure which I am causing it to incur? Will this bit of investigation, that bit of research, these tabulations, those data gathered and interpreted yield a return in financial savings or increased efficiency in the system which can be made apparent to the school administration and the public in general? There are scores and scores of problems which have not been solved. There are many, many questions to which we must now answer that we do not know. There are all sorts of interesting things which may be found out about schools and education. But it is quite probable that in the present period of financial depression and restless questioning, the public—the paying public—is but little interested in pure research, in the answer to idle questions, or in the solution of problems which do not appeal to it and which you may not have demonstrated to be of real vital worth.

I do not believe that the public insists or will insist that the evaluation of all educational effort shall be expressed in economic terms. I am quite sure it is interested in spiritual values. But I am also quite sure that this same public is interested and will more or less definitely insist upon being shown returns for the investment which is made in terms which it can appreciate and which it will approve.

One of the members has written to your secretary for his opinion on the following points. Cannot the membership of the association contribute and thus enable him to answer with a composite judgment instead of an individual one? The questions are as follows:

1. From the standpoint of state administration, would it be advisable to have the same age definitions for intelligence test records and for reporting age-grade census data?

2. If one standard is fixed, is it better to consider an eight-year-old norm as including all the children of ages 8 years 0 months to 8 years 11 months inclusive or should we consider an eight-year-old as being anywhere between 7 years 7 months and 8 years 6 months inclusive?

3. In establishing age norms on either of these bases, is it better to group children in year groups or in half-year groups?

Dr. Franzen in charge of the research work in the Des Moines, Iowa schools sends in a twenty-point program of his work there. There is space for but a mere statement of the main points.

1. City-wide measurement of intelligence
2. City-wide measurement of arithmetic
3. Intensive measurement in three schools in terms of Educational Quotient
4. Measurement of product in high schools
5. Measurement of intelligence of incoming freshman class in one High School
6. City-wide measurement of handwriting
7. Formulation, etc. of test for incoming teachers
8. Construction of objective composition scale
9. Construction of scales, typewriting test
10. Construction of intelligence test for kindergarten children
11. An experiment in the ethics of school children
12. Measurement of oral and silent reading in grades I, II, and III
13. An experiment in the pedagogy of shorthand
14. Comparative statistics on athletic records
15. Course in statistics and measurements for principals and supervisors
16. Help in diagnosis of difficulties
17. Meeting with teachers whenever information or advice will make better teachers
18. Conferences with supervisors
19. Determination of a workable and reasonable salary schedule
20. Computation and arrangement of tables and of technic which will increase efficiency of office routine.

Dean Russell of the State University of Iowa has called a week's conference of superintendents who had not planned to attend summer school to meet at the

University about the middle of July to work intensively on problems of school finance. About a dozen men volunteered at the recent Conference on School Supervision to attend and work with him on this problem.

Several subcommittees of the standardization committee reported at the closed meeting on February 26, 1921. The report of the subcommittee on statistical methods was read by Dr. Rugg in the absence of the writer Dr. Truman L. Kelley. This report follows.

REPORT OF THE SUBCOMMITTEE ON STATISTICAL METHODS OF THE STANDARDIZATION COMMITTEE (TENTATIVE)

There is a variety of procedure in the treatment of test and measurement data. This is unfortunate from the standpoint of ready interpretation of published studies, but it is indicative of the virility of the movement. The number of workers each believing in the superior merit of his method of presentation makes it undesirable, without more information than is now available to your committee to take issue with any of the methods employed, except as implied in proposals (1) and (2). For the reasons stated and because statistical procedure in psychology and education should conform with that already established in the older biological and physical sciences your committee thinks it desirable to offer general rather than specific proposals:

1. Any worker presenting a new procedure should definitely recognize that it is impossible to prove the superiority of his method by reporting data interpreted by means of his method alone. Superiority is a relative matter and it is necessary to compare a method with alternative methods before its superiority can be established.

2. Any worker presenting a new procedure should definitely expect that the burden of proof lies with him. This implies that he should consider himself responsible for proving: (a) that his method is more reliable in the sense that the probable errors of results obtained by it are less than those obtained by the more standard methods. Proving this would give the method prior claim to excellence and it would then devolve upon other claimants to establish the superiority of their methods. Or, (b) that the method has advantages of interpretation or of expedition that more than compensate for its lessened reliability. This implies that the amount of this lessening in reliability is established and made known by the worker.

The following suggestions as to various measures are offered:

3. Of the measures of central tendency the arithmetic average or mean is easily understood, easily calculated, and usually for such distributions as are found in educational research work, has a smaller probable error than the median or mode. The median, as proved by educational studies of the last decade, is calculated in many ways. Its probable error is almost always greater than that of the mean, and has seldom been reported. Its brevity (not its accuracy) of calculation and ease of interpretation recommend it but in case these are not prime considerations your committee recommends the use of the mean in place of the median or mode as a measure of central tendency.

4. Of the measures of dispersion the standard deviation usually has, for such distributions as are found in educational studies, a smaller proportionate probable error than the average deviation or the quartile deviation. Its general use in scientific presentations as a measure of dispersion is recommended. For popular presentation

0.6745 times the standard deviation, or the probable error, is recommended. If necessary considerations have lead to the use of the median instead of the mean as a measure of central tendency, it is recommended that a measure of dispersion based upon percentiles, e. g. the quartile deviation, the 10–90 percentile range, etc., be used as the measure of dispersion.

5. Of the measures of correlation, the Galton-Pearson product moment coefficient of correlation is well established, having, in the case of rectilinear relationships, the smallest probable error of any of the coefficients of correlation. The Spearman ρ coefficient based upon the squares of the differences in rank holds a similar position in the case of ranked data. These two coefficients are recommended because of their known properties and particularly because their probable errors are available. A measure of relationship for which a formula giving the probable error is not available should be considered as merely in the experimental stage.

Many devices have been proposed for the measurement of non-rectilinear relationships. No recommendation with reference to them is offered.

6. As a general statement of policy to be followed in educational research your committee would recommend that the most reliable of several available measures (such, for example, as measures of central tendency, of dispersion, of correlation, of frequency, etc.) be used, and that when one departs from this rule he should feel it incumbent upon himself to justify his procedure upon substantial grounds, such as the necessity for using a less time consuming method, greater simplicity of interpretation, etc.

TRUMAN L. KELLEY
Chairman, Subcommittee on Statistical Methods

At the closed meeting of the association at Atlantic City on February 26, 1921, Mr. Courtis, chairman of the standardization committee appointed at the Cleveland meeting, presented material gathered from the questionnaire which he had sent out to the members. On the following Thursday he made the report of the committee and the report was ordered printed. We present it herewith.

REPORT OF THE STANDARDIZATION COMMITTEE

Tabulation of the returns from the questionnaire sent to members revealed two outstanding facts: (1) a practically unanimous sentiment in favor of the publication of an official list of terms, procedures, etc. and (2) a fear that the attempted standardization may be premature. Therefore the statements following have been prepared as the recommendation of the committee, but it is suggested that the list be printed in the JOURNAL OF EDUCATIONAL RESEARCH as tentative standardizations for trial purposes only, final approval to be deferred until the annual meeting in 1922. During the interval use of the approved forms is to be experimental and voluntary, but members preparing material for publication are asked to try to conform to these regulations.

1. Members of the association preparing material for publication, and wishing to use alternative forms or definitions are asked to give reasons or explanations in footnotes.

2. A test composed of elements of uniform difficulty, or of several cycles of uniform difficulty, and used to determine the rate at which the work is done, should be called Rate Tests. (*Illustrations:* Cleveland Arithmetic Tests, Starch's Reading Tests, Courtis Writing Tests, etc.)

3. The term "scale" should be applied only to tests or material graded in difficulty, or quality, and used to measure degree of difficulty or quality.

4. A test composed of elements graded in difficulty and used to determine the most difficult test material the subject can handle successfully, under the prescribed conditions, should be called a performance scale. (*Illustrations:* Trabue's Completion Test, Language Scales, Gray's Oral Reading Scale, Thorndike's Reading Scale.)

5. Whenever the objective product of a pupil's test is measured as to quality by comparison with samples of known value, the collection of samples of known value should be called a Product Scale. (*Illustrations:* Thorndike Handwriting Scale, Hillegas Composition Scale, Rugg's Lettering Scale.)

6. In connection with both rate tests and performance scales the score obtained is frequently evaluated in terms of the results of giving the test to age or grade groups. This evaluation consists in placing the score upon a scale of difficulty. Where this form of scaling has been made possible, the testing instrument itself is often called a scale. (*Illustrations:* Burgess Reading Scale, Ayres Spelling Scale, Hahn-Lackey Geography Scale, etc.) At this time the committee is not able to recommend a suitable name for scales of this type, or for other tests and scales which are mixtures or combinations of the simpler forms named above.

7. The term "standard" should be applied to scores as an adjective, and only when the scores are set up as goals or objectives of teaching effort. If adopted by an authoritative organization, they may be called "standard scores" without qualification but scores set up as goals by an author should be designated as "author's standard scores." A score representing an actual performance of a group should be called a "norm." Whenever norms are published they should be accompanied by statements of the number of cases, upon which each norm is based and of the statistical procedure by which it was derived.

8. The term "average" should be used only in its generic sense to include any and all measures of central tendency. The term "arithmetic mean" should be used as the name of the sum of a series of measures divided by the number of measures. (*Illustration of correct use of term:* Averages in common use are the mean, the median, and the mode.)

9. The term "mean" when used alone should be taken to connote the "arithmetic" mean. All other types of means should be used with descriptive adjectives, as harmonic mean, geometric mean, etc.

10. There is need to distinguish between measures of central tendency derived from actual scores and those derived from frequency distributions. For instance, the new term "midscore" is suggested as a name to be applied to the middle most measure (or mean of the two middle measures when there are two), while the term "median" should be reserved for the value derived from a frequency distribution by interpolation. (See Rugg's *Statistical Methods Applied to Education*, pages 103-14.)

11. The term "Performance" should be used to connote the score obtained in a test. When, however, there is a desire to convey the idea of comparison with norms or standards the term "achievement" should be used. The two terms are, thus, not precisely synonymous.

12. The score of an individual in a test should not be considered as a measure of his ability but as a record of his performance under the particular conditions under which the test was given. Judgments as to ability are always inferences from performance and the greater the number of trials or tests from which ability is inferred, the greater the probable correctness of the inference.

13. The term "capacity" should be restricted so that it may connote just one thing: namely, inborn, or undeveloped possibilities of behavior.

14. Ability should be regarded as the product of experience acting on capacity.

15. The movement for the use of measurement in education has had for its goal the attainment of uniformity and certainty of interpretation. Great advances have already been made but still greater advances are possible. To aid in bringing about improvement, makers of tests should be careful to investigate the effect of every factor involved in standardization, conforming so far as possible to the law of the single variable. In general the process of standardization of a test will involve the following operations (although not necessarily in the order given):

A. Preparation and selection of test material
B. Experimental organization of test and instructions for giving the test.
C. Trial of tentative test to determine value of elements, gross validity, reliability, and optimum conditions of giving, scoring, etc.
D. Final organization of test.
E. Final formulation of conditions under which test is to be given, scored, tabulated, and interpreted.
F. Official determination of validity
G. Official determination of reliability
H. Official determination of norms.

The association accordingly invites makers of new tests to file with its standardization committee for publication data which throw light upon

A. What a test measures.
B. How reliably it measures.
C. The principles of formulae of construction.
D. Statements of conditions for giving, scoring, and tabulating the test and for interpreting the results.
E. Norms.

16. Members of the association and others are invited to send to the committee suggestions for the modification of the above, or other definitions or practices which should be standardized.

Problems

Two of the most important types of problems in measurement are those connected with the determination of what a test measures, and of how consistently it measures. The first should be called the problem of validity, the second, the problem of reliability.

Members are urged to devise and publish means of determining the relation between the scores made in a test and other measures of the same ability: in other words, to try to solve the problem of determining the validity of a test.

Members are also urged to study carefully the relations between the scores made in a test and the scores made in other trials of the same or equivalent editions of a test (the problem of reliability).

Norms should probably be set only after a careful study of the effects of variation of each of the factors of age, grade, time of year, heredity, maturity, training, and social status. As rapidly as proves feasible correction tables for variations in each of these factors should be worked out and published.

Respectfully submitted, B. R. Buckingham, W. A. McCall, A. S. Otis, H. O. Rugg, M. R. Trabue, S. A. Courtis, *Chairman, Special Standardization Committee.*